The Role of *Arabidopsis thaliana* Mismatch Repair Proteins in Meiotic Recombination

Alexander R. Blackwell

Department of Plant Sciences
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

Gonville & Caius College

January 2019

Declaration of originality

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Biology Degree Committee.

Alexander R. Blackwell

January 2019

The Role of *Arabidopsis thaliana* Mismatch Repair Proteins in Meiotic Recombination

Alexander R. Blackwell

Abstract:

Meiosis is a conserved eukaryotic cell division that increases genetic diversity in progeny. During meiosis, homologous chromosomes pair and undergo reciprocal exchange, called crossover. Meiotic recombination initiates from DNA double strand breaks, which are repaired using either sister or homologous chromatids as templates. When meiosis occurs in heterozygous (hybrid) organisms, interactions between homologous chromosomes have the potential to generate mismatched DNA structures. Several proteins with roles in DNA mismatch repair (MMR) are known to influence meiotic recombination in several eukaryotes. In *Arabidopsis thaliana* this includes three MSH heterodimers (MSH2-MSH3, MSH2-MSH6 and MSH2-MSH7) that recognise mismatched nucleotides and have demonstrated roles in repressing meiotic crossovers in hybrid plants.

To further investigate the meiotic function of MMR genome-wide, I generated a series of *msh2* mutant introgression lines in three different genetic backgrounds – Ct-1, L*er*-0 and CLC – which have varying patterns and levels of polymorphism. Consistent with the function of MSH2 as a hybrid-specific anti-recombinase, I observed significant crossover increases in the chromosome arms in all three hybrid *msh2* mutants. However, I also found evidence for accession and region specific effects of *msh2* on crossover frequency. For example, crossovers appeared to decrease over centromere proximal regions in *msh2* compared to wild type. A genotyping-by-sequencing experiment was performed to generate genome-wide crossover maps in two *Arabidopsis* hybrids, with and without MSH2 function. This revealed that total crossover number remained unchanged in the MMR-deficient hybrids, whilst crossovers redistributed into regions of reduced polymorphism density. This relationship was counter to my expectation that MMR would repress crossovers most strongly in divergent regions. However, this relationship was observed across varying physical scales, from 1–100 kilobases. This confirms a positive relationship between polymorphism and meiotic crossovers in *Arabidopsis* hybrids, and reveals a novel role for MSH2 in mediating this effect.

In addition to the investigation of MSH2 in meiotic recombination, I present an analysis of the genome-wide distribution of MutS homolog 4 (MSH4). MSH4, and its binding partner MutS homolog 5, evolved from their ancestral role in MMR and now function exclusively to promote meiotic crossovers in the ZMM pathway. I present a genome-wide chromatin-immunoprecipitation-sequencing analysis of the binding profile of MSH4, and analyse its distribution at varying scales. This has revealed novel relationships between MSH4, the meiotic cohesin subunit REC8, and features of the chromatin and recombination landscapes. Together, these results advance our understanding of meiotic recombination in plants, and raise further questions about the regulation of crossovers in eukaryotes more broadly.

I would like to dedicate this thesis to the memory of Dr Ian Moore (1964-2018), an inspiring teacher who I never thanked enough.

Acknowledgements

Firstly, I would like to thank The Gatsby Charitable Foundation for providing me with a Sainsbury Ph.D Studentship, and thus the opportunity to continue studying what I love. Gatsby have provided me with invaluable support over nearly six years, which will unquestionably shape my life for much longer. In particular, I would like to thank the Gatsby Plant Science Advisors – Liam Dolan (University of Oxford), Jane Langdale (University of Oxford), Ottoline Leyser (The Sainsbury Laboratory, Cambridge), and Nick Talbot (The Sainsbury Laboratory, Norwich) – who have provided me with advice and guidance in innumerable ways.

My Ph.D would not have been possible without the vision and guidance of Ian Henderson, whose constancy and motivation has been awe inspiring. It's been quite a journey over the last four years, and Ian's continual enthusiasm and energy was mostly a great help! I am especially grateful for the opportunities he provided for me to present my work, and for the flexibility, support and advice he offered when my daughter was born mid-Ph.D. It has been a wonderful and thoroughly educational experience to work together.

I would also like to thank the postdocs in the group, past and present, for sharing ideas, offering advice, and providing interesting company: Piotr Ziółkowski, Natasha Yelina, Kyuha Choi, Xiaohui Zhao, Sabrina Gonzalez-Jorge, Heïdi Serra, Christophe Lambing, Joiselle Fernandes, Wei Jiang, and Andrew Tock. Christophe has been a mentor throughout my time in the lab, and helped me get up and running. Thank you for your patience and inspirational work ethic.

I would like to acknowledge all of the graduate students in both Ian Henderson's group and David Baulcombe's groups, past and present, who I had the good fortune to overlap with: Charles Underwood, Alex Canto-Pastor, Luke Braidwood, Claire Agius, Patrick Diaz, Emma Lawrence, Quentin Badolle, Divyashree Nageswaran, Pallas Kuo, Quentin Gouil, Daniel Holland, Piotr Włodzimierz, and Sarah Garland. Together, you made my time in Cambridge more enriching and thought provoking.

I would like to thank Mel Steer, Stephanie Topp, James Barlow, and Pawel Baster for technical and horticultural assistance. Avraham Levy (The Weizmann Institute), Gregory Copenhaver (University of North Carolina Chapel Hill) and Scott Poethig (University of Pennsylvania) kindly shared Arabidopsis resources. I would also like to thank Kyuha Choi (Pohang University of Science & Technology) for sharing his plasmid constructs and ChIP-seq protocols, and for answering my many questions with such patience and wisdom.

Outside of the laboratory, I am was fortunate to be part of the Old Science Society, together with Om Patange, Michael Kosicki, Patrick Diaz, Andrew Mock, Quentin Badolle, Bruno

Martins and Giulia Arsuffi. Thank you for the excellent and wide-ranging discussions, and for remembering the bigger picture. I hope we find a way to keep this going.

To my partner, Milja Fenger: thank you for being part of my life over the last five years, for your commitment to our life together, and for your support, patience and self-sacrifice over the last six months. It certainly feels like one era is ending, as we step into the next. To my daughter, Jesseke: thank you for joining our family, for being a source of joy and belachelijkheid, and for reminding me of what's really important. Also, thank you for arriving at a time that forced me to become more organised and effective. To my parents: I cannot thank you enough for your love and support over the 26 years of my life, and for offering the security that has enabled me to progress with my academic work.

Finally, I would like to thank Matt Neale (University of Sussex) and Julian Sale (MRC LMB, Cambridge) in advance, for agreeing to be my examiners and for taking the time to read this thesis.

Table of Contents

Table of Figures & Tables

List of Abbreviations

A	Adenine
Arabidopsis	*Arabidopsis thaliana*
bp	Base pair
C	Cytosine
CAPS	Cleaved amplified polymorphic sequence
CG	DNA cytosine methylation in CG contexts
CHG	DNA cytosine methylation in CHG contexts (H = A, C or T)
CHH	DNA cytosine methylation in CHH contexts (H = A, C or T)
ChIP-seq	Chromatin immunoprecipitation-sequencing
CLC	Cvi-0, L*er*-0, Col-0
cM	Centimorgan
CTAB	Cetyl trimethylammonium bromide
CTL	Col-0 traffic line
dHJ	Double Holliday junction
DNA	Deoxyribonucleic acid
D-loop	Displacement loop
DSB	Double strand break
EMS	Ethyl methanesulfonate
FDR	False discovery rate
FTL	Fluorescent tagged line
G	Guanine
GA	Gibson assembly
GBS	Genotyping-by-sequencing
gDNA	Genomic DNA
GFP	Green fluorescent protein
GUS	β-glucuronidase
H3K27me1	Monomethylation of histone 3 lysine 27
H3K27me3	Trimethylation of histone 3 lysine 27
H3K36me3	Trimethylation of histone 3 lysine 36
H3K4me3	Trimethylation of histone 3 lysine 4
H3K9me2	Dimethylation of histone 3 lysine 9
HA	Hemagglutinin
hDNA	Heteroduplex DNA

HJ	Holliday junction
HR	Homologous recombination
Indel	Insertion or deletion
JM	Joint molecule
kb	Kilobase
LTR	Long terminal repeats
Mb	Megabase
MDR	Methyl-directed repair
MMR	Mismatch repair
NASC	Nottingham Arabidopsis Stock Centre
nt	Nucleotide
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PSB	Pollen Sorting Buffer
RFP	Red fluorescent protein
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing
RPKM	Reads per kilobase of transcript per million mapped reads
rpm	Rotations per minute
SEI	Single-end invasion
SNP	Single nucleotide polymorphism
SSA	Synthesis-dependent strand annealing
ssDNA	Single-stranded DNA
SSLP	Simple sequence length polymorphism
T	thymine
T-DNA	Transfer DNA
TE	Transposable element
TSS	Transcriptional start site
TTS	Transcriptional termination site

Chapter 1 – Introduction – Meiosis, mismatch repair, and recombination

In this chapter, I will firstly give an overview of meiotic recombination. Secondly, I provide an overview of DNA mismatch repair (MMR) in both bacteria and eukaryotes, highlighting the evolutionary conservation of MMR proteins in plants. Thirdly, I will examine the literature relating to MMR's role in both somatic and meiotic recombination, particularly in the context of sequence divergence between recombining substrates. Finally, I will outline the aims and objectives of this research project. Throughout this introduction, a focus will be placed on the model system *Arabidopsis thaliana* (hereafter Arabidopsis)

1.1 Meiotic recombination

1.1.1 An overview of meiosis

Sexual reproduction is a fundamental biological process in eukaryotes, which is believed to have evolved to promote genetic variation in progeny (Otto & Gerstein, 2006). Central to this process is meiosis, a highly conserved form of the cell cycle which produces four haploid gametes from a single diploid meiocyte (Ohkura, 2015). It functions to reduce ploidy by half, and re-shuffles genetic variation between paternally and maternally inherited chromosomes (Ohkura, 2015). In contrast to mitosis, high levels of DNA double strand breaks (DSBs) are induced after completion of meiotic S-phase, which promotes interhomolog recombination (Lam & Keeney, 2014). A subset of these interhomolog interactions progress to form reciprocal exchanges between chromosomes called crossovers, which manifest cytologically as chiasmata (Holliday, 1964; Hunter, 2015). Crossovers are required, together with sister chromatid cohesion, for the alignment of homologous chromosome pairs on the spindle axis (Hunter, 2015). Crossover resolution at metaphase I allows segregation of the recombined chromosomes to opposite cell poles (fig. 1) (Hunter, 2015).

In contrast to mitosis, kinetochores, the macromolecular complexes that mediate interactions between chromosomes and spindle microtubules, must be mono-orientated for replicated sister chromatids, in order for segregation of homologs at meiosis I (Hauf & Watanabe, 2004). Centromere mono-orientation requires the protection of centromeric cohesion during meiosis I, which is in contrast to mitosis (Ohkura, 2015). In the majority of eukaryotes examined, accurate chromosome segregation during meiosis requires at least one 'obligate' crossover per homologue pair (Martini et al., 2006; Jones & Franklin, 2006). In the absence of an intervening round of DNA replication, the two chromosome sets proceed to segregate sister

chromatids via loss of centromeric cohesion at metaphase II, producing four haploid gametes (Ohkura, 2015) (fig. 1).

Mechanisms of meiotic recombination have been intensively studied in plants, and there are now over 90 genes with characterised roles in meiosis (Mercier et al., 2015). Although both meiosis I and meiosis II are divided into prophase, metaphase, telophase and anaphase stages, the following introduction will focus exclusively on prophase I, a stage that is temporally extended in comparison to mitotic prophase and is the phase of meiotic recombination (Ross et al., 1996; Armstrong, 2013). Meiotic prophase I is itself divided into several distinct sub-stages. In leptotene, chromosome condensation begins and the chromosomes form thread-like filaments (Ma et al., 2006). Chromosomes then interact to form homologous pairs, a process dependent on sequence homology in most eukaryotes. Pairing leads to the establishment of a proteinaceous axis termed the synaptonemal complex, and a stable juxtaposition of homologous chromosome pairs termed synapsis (Ma et al., 2006). When synapsis is partially completed, meiocytes enter a sub-stage termed zygotene. Upon completion of synapsis, the chromosomes appear as thick threads, a sub-stage termed pachytene. After pachytene the synaptonemal complex disassembles, causing the homologous chromosomes pairs to separate, whilst remaining associated at sites of meiotic crossover. The chromosome pairs then re-condense at the diakinesis sub-stage, the final sub-stage of meiotic prophase I, where the chromosomes become visible as distinct pairs (Ma et al., 2006). In Arabidopsis, the male meiotic cell cycle has a duration of 33 hours, with a large proportion of this being occupied by zygotene/pachytene (Armstrong et al., 2003).

1.1.2 Formation of meiotic DSBs

Meiotic recombination initiates with the formation of DNA DSBs. Consistent with the ‘DSB repair model’, proposed by Szostak et al. (1983), Spo11 induces meiotic DSBs (fig. 2A) (Keeney et al., 1997; Keeney & Neale, 2006). This protein is highly conserved, and has homology with archaeal topoisomerase VI (Keeney et al., 1997; Keeney & Neale, 2006). A minimum of nine meiotic DSB cofactors have been identified in budding yeast: Mre11, Rad50, Xrs2, Ski8, Rec102, Rec104, Rec114, Mei4, and Mer2 (Keeney & Neale, 2006). Plant orthologues of these factors exist, including PRD2 (Mei4), but these factors are generally greatly diverged at the sequence level (De Muyt et al., 2009). For example, PRD1 is a homolog of human Mei1, but has only 22% amino acid identity. Other cofactors are likely plant-specific, such as the DSB-regulating PRD3 and DFO in Arabidopsis (De Muyt et al., 2007; De Muyt et al., 2009; Zhang et al., 2012).

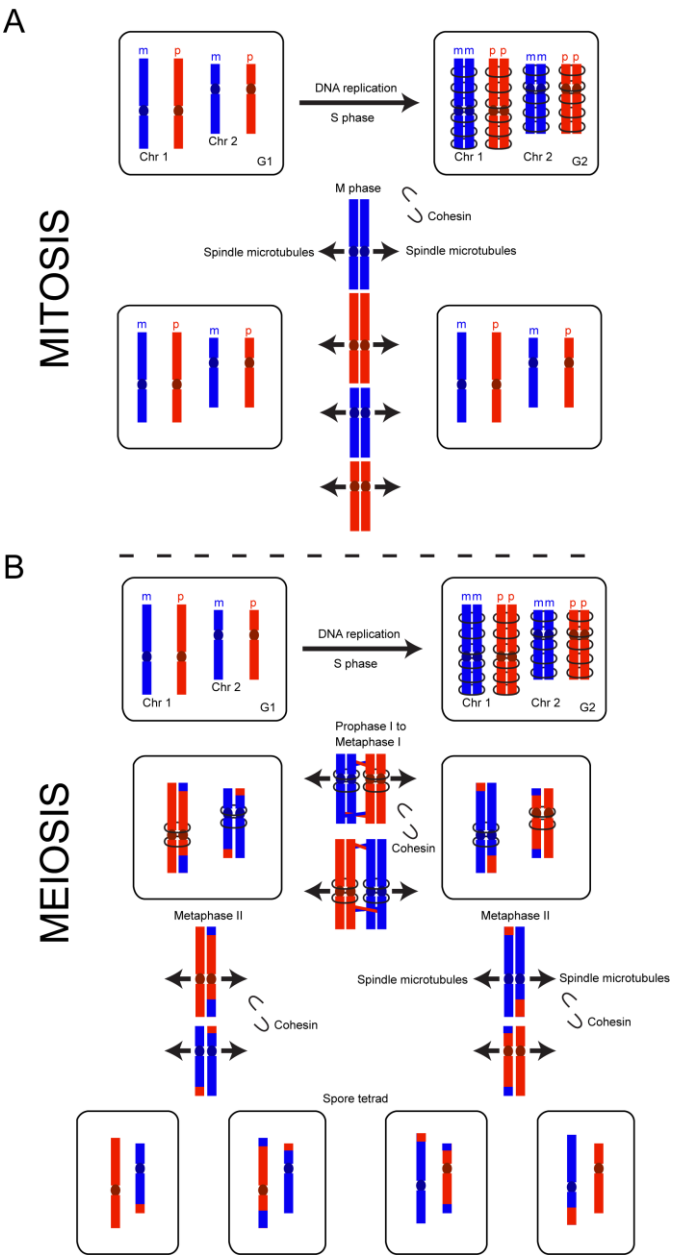


Figure 1. Schematic representation of the mitotic and meiotic cell cycles.

(A) A schematic representation of mitosis. Following G1, maternal (m) and paternal (p) chromosomes replicate, and are enclosed by cohesin. After G2, chromosome pairs align and their kinetochores bi-orientate with the mitotic spindle microtubules. Cohesin is cleaved at metaphase, and chromatids segregate into two daughter cells. (B) A schematic representation of meiosis. Like mitosis, maternal and paternal chromosomes replicate at S-phase and are enclosed by cohesin. However, meiotic recombination occurs and associates the homologous chromosome pairs, promoted by synapsis and crossover formation (red/blue exchanges). Sister kinetochores are mono-orientated, and chromosome pairs segregate to opposite poles at metaphase I after crossover resolution. Meiotic cohesin is retained in proximity to the centromere to maintain sister association. With no intervening round of DNA replication, remaining cohesin is cleaved and sister chromatids segregate into the four products of meiosis at metaphase II.

The Arabidopsis genome encodes three *SPO11* orthologs, *SPO11-1*, *SPO11-2* and *SPO11-3*, although the third is not required for meiosis (Grelon et al., 2001; Stacey et al., 2006; Hartung et al., 2007). Mutants in either *SPO11-1* or *SPO11-2* fail to initiate meiotic recombination, causing defects in chromosome pairing and synapsis, and are thought to function as a SPO11-1/SPO11-2 heterodimer (Stacey et al., 2006; Hartung et al., 2007). The requirement of DSB formation for synapsis in Arabidopsis is consistent with mouse and many fungi, but contrasts with *Drosophila melanogaster* (hereafter Drosophila) and *Caenorhabditis elegans*, where synapsis is independent of recombination (Dernburg et al., 1998; Baudat et al., 2000, Romanienko and Camerini-Otero, 2000; Comeron et al., 2012). A second archaea-related ortholog, a sub-unit B-like topoisomerase (MTOPVIB), was found to function in a complex with SPO11-1 and SPO11-2, in both Arabidopsis and mouse (Robert et al., 2016; Vrielynck et al., 2016). In Arabidopsis, MTOPVIB interacts with SPO11-1 and SPO11-2, and the *topoVIB* mutant phenocopies *spo11-1* and *spo11-2* mutants (Vrielynck et al., 2016). These findings indicate that meiotic DSBs are induced by a conserved topoisomerase VI-related heterotetramer (Robert et al., 2016; Vrielynck et al., 2016).

Quantifications of DMC1 foci from Arabidopsis meiotic chromosome spreads estimate that 150-250 meiotic DSBs form during meiosis, despite only ~10 crossovers occurring per meiosis (Chelysheva et al., 2007; Sanchez-Moran et al., 2007). This disparity between the number of recombination initiation foci and final crossover number indicates that only a subset of meiotic DSBs can become designated as crossovers. A recent study demonstrated the dosage dependence of *SPO11-1*, where hypomorphic *spo11-1* transgenic lines were observed to have ~40% reductions in the DSB markers γH2A.X and RAD51 (Xue et al., 2018). However, this DSB reduction was associated with a reduction of ~1.5 crossovers per meiosis, implying that homeostatic mechanisms regulate global crossover number in Arabidopsis, for example by limiting the activity of anti-crossover mechanisms (Xue et al., 2018). This finding is consistent with previous observations of crossover homeostasis in budding yeast and mouse (Martini et al 2006; Cole et al., 2012).

1.1.3 Processing of meiotic DSBs

During DSB formation, both SPO11 subunits are covalently bound to 5′ DNA ends, via a conserved tyrosine residue (Keeney et al., 1997; Lam & Keeney, 2014). A strand incision is then made on each DNA strand, one on each side of the DSB (Neale et al., 2002). Endonucleolytic processing, by the conserved Mre11-Rad50-Xrs2/Nbs1 (MRX/N) complex and Com1 (Sae2), then resects towards SPO11 (Neale et al., 2005; Garcia et al., 2011; Shibata et al., 2013). Sae2 is known to promote MRX/N activity, via phosphorylation-induced tetramerisation of Sae2 and its interaction with Rad50 (Cannavo et al., 2018). The Arabidopsis

ortholog of *Com1*/*Sae2* also plays an essential role in meiotic DSB resection, and *com1*/*sae2* mutants fail to form RAD51 foci and accumulate SPO11-1 foci, potentially indicating a failure to process meiotic DSBs (Uanschou et al., 2007). Further DSB processing is performed by Exo1, which produces an extensive region of 3′ ssDNA on either side of the DSB (fig. 2B) (Garcia et al., 2011; Shibata et al., 2013). Together these observation indicate that meiotic DSB processing is likely conserved between Arabidopsis and other eukaryotic model systems.

Bidirectional DSB resection causes the release of SPO11-1-oligonucleotide complexes covalently bound to short polynucleotides, 30-40 bases in length in Arabidopsis (Choi et al., 2018), which has enabled DSB mapping in budding yeast, fission yeast, mouse and Arabidopsis (Pan et al., 2011; Fowler et al., 2014; Lange et al., 2016; Choi et al., 2018). Exposed 3′ ssDNA ends are then bound by replication protein A (RPA) (Osman et al., 2009; Ribeiro et al., 2016). Although yeast and mammals encode all three RPA subunits (RPA1, RPA2 and RPA3) as single-copy genes, plants encode multiple copies of each subunit: five *RPA1* genes, two *RPA2* genes and two *RPA3* genes (Shultz et al., 2007; Aklilu et al., 2014). *RPA1a* is specifically required for Class I crossovers (see section 1.1.7), as the *rpa1a* mutation causes a 60% reduction in chiasma frequency and reduced fertility (Osman et al., 2009). Hence, although the process of DSB formation appears broadly conserved, several components involved in DSB processing have undergone gene duplications in Arabidopsis, as evidenced by the enlarged *RPA* and *SPO11* gene families (Shultz et al., 2007; Aklilu et al., 2014; Stacey et al., 2006; Hartung et al., 2007).

1.1.4 Interhomolog strand invasion

After DSB processing, strand invasion and homolog search are mediated by nucleoprotein filaments (Da Ines et al., 2013). In Arabidopsis, the recombinases *RAD51*, *RAD51C* and *XRCC3* are required for filament formation and successful somatic and meiotic recombination, as their deletion causes a *spo11-1* dependent meiotic chromosome fragmentation (Li et al., 2004a; Li et al., 2005), indicating that RecA family proteins are required to mediate meiotic DSB repair. The RecA family member DMC1 plays a meiosis specific role in mediating interhomolog strand invasion (Bishop et al., 1992; Yoshida et al., 1998; Da Ines et al., 2013). In Arabidopsis *dmc1*, DSBs are repaired using sister chromatids and univalents segregate at meiosis I (Couteau et al., 1999). In contrast, *rad51* causes meiotic chromosome fragmentation and complete sterility (Da Ines et al., 2013). Together, these data indicate that DMC1 promotes interhomolog recombination in meiosis, with RAD51 supporting strand invasion and mediating intersister recombination in the absence of DMC1.

1.1.5 DSB-associated DNA synthesis

Strand invasion forms a displacement loop (D-loop), where one strand of the invaded homolog is displaced, and DNA synthesis is primed at the end of the invading strand (fig. 2C,D) (Wang & Copenhaver, 2018). In Arabidopsis, studies of crossover and non-crossover associated gene conversions were used to estimate tract lengths occurring with each type of repair (Lu et al., 2012; Sun et al., 2012; Wijnker et al., 2013). Gene conversion is the process whereby DNA synthesis occurs over a heterozygous site following interhomolog strand invasion, following which the synthesized DNA strand can re-anneal to the parental chromosome. This results in a mismatch at the heterozygous site that can be repaired via mismatch repair (MMR), leading to conversion of one allele to the other and thus causing 3:1 inheritance of an allele through meiosis (Holliday, 1964; Wijnker et al., 2013). Sequencing of Arabidopsis tetrads estimated that crossover associated conversion tracts were longer than non-crossover associated tracts (~400 bp versus ~25-50 bp, respectively) (Wijnker et al., 2013). This observation is consistent with observations of increased DNA synthesis at sites of meiotic crossover in budding yeast (Terasawa et al., 2007).

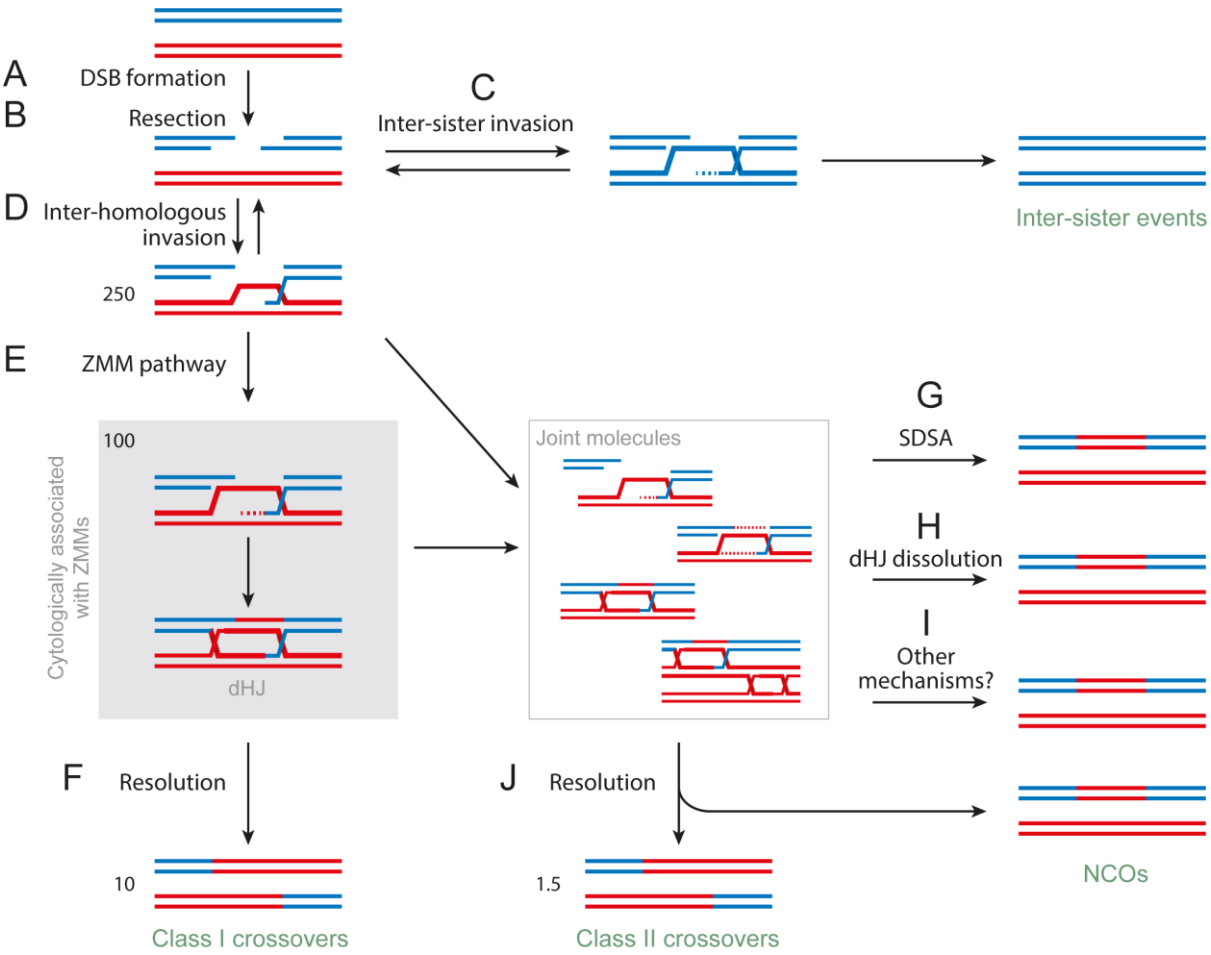


Figure 2. Schematic representation of meiotic recombination. Adapted from Mercier et al. (2015).

A schematic representation of meiotic recombination pathways. Meiotic recombination initiates with the formation (A) and resection (B) of DSBs, which can then invade either homologous (D) or sister chromatids (C) leading to a D-loop. After DNA synthesis (dotted lines), ZMM proteins then promote dHJ formation (E) which can be resolved as Class I crossovers (F). Alternatively, interhomolog D-loops can mature into various joint molecule (JM) intermediates, which can form non-crossovers via SDSA (G), dHJ dissolution (H), or additional unknown mechanisms (I). A second, ZMM-independent crossover resolution pathway forms Class II crossovers (J). Arrows indicate reversible or non-reversible steps. Red and blue lines indicate strands of DNA. The estimated numbers of several DNA intermediates in Arabidopsis are indicated.

1.1.6 Recombination and the meiotic chromosome axis

Meiotic recombination takes places in the context of a specialised chromosome axis (fig. 3A) (Zickler & Kleckner, 1999). Chromosomes are connected to a linear proteinaceous axis that serves as a scaffold to recruit key meiotic proteins (Börner et al., 2004; Kleckner, 2006; Sanchez-Moran et al., 2007; Ferdous et al., 2012). After meiotic S-phase, sister chromatids are held together by cohesin complexes, which become associated with the axial element (Zickler and Kleckner, 1999). Sister chromatids are co-aligned in linear arrays of chromatin loops, with the base of the loops tethered to the axial element (Blat et al., 2002; Storlazzi et al. 2010; Panizza et al. 2011). In Arabidopsis, the axis is comprised of HORMA domain proteins such as ASY1, and its partner ASY3, which stimulate interhomolog recombination and crossover formation (Armstrong et al., 2002; Ferdous et al., 2012). In Arabidopsis, a further axis protein ASY4 interacts with ASY1 and ASY3, and ASY3/ASY4 are potentially the functional orthologs of the mammalian SYCP2/SYCP3 axial components (Yang et al., 2006; Syrjänen et al., 2014; Chambon et al., 2018). As prophase I progresses, the HORMA domain axis proteins are depleted and synaptonemal complex components, such as ZYP1 are loaded, with synapsis completing at pachytene (Higgins et al., 2005; Ferdous et al., 2012; Lambing et al., 2015). The meiotic chromosome axis is required for synapsis, where homologous chromosome pairs are brought into close association via the formation of a central element (Higgins et al., 2005). Hence, meiotic recombination takes place in the context of a complex protein axis.

Interestingly, whilst meiotic recombination in many eukaryotes depends on the underlying synaptonemal complex, the process of synapsis also depends on recombination (Kerzendorfer et al., 2006; Zickler & Kleckner, 2015). For example, in Arabidopsis, synapsis depends on the formation of recombination intermediates, as the synaptonemal complex is absent in *spo11-1* or *dmc1* mutants (Grelon et al., 2001; Da Ines et al., 2013). In relation to the meiotic

chromosome axis, Spo11 and DSB accessory proteins, such as Mei4, Mer2 and Rec114, stably interact with chromosome axis sequences in budding yeast and mouse, where recombination occurs (Pan et al., 2011; Baudat et al., 2013; Panizza et al., 2011; Stanzione et al., 2016). For instance, direct interactions were observed between the H3K4me3 binding protein Spp1 and the meiotic axis component Mer2 (fig. 3B) (Acquaviva et al., 2013; Sommermeyer et al., 2013). This is consistent with the observation that budding yeast and Arabidopsis DSB hotspots are associated with H3K4me3-modified nucleosomes (Pan et al., 2011; Choi et al., 2018). These observations are consistent with a 'tethered loop/axis model', where chromatin loops are tethered to the axis via REC8-cohesin during meiotic DSB induction and crossover repair (Blat et al., 2002; Pan et al., 2011).

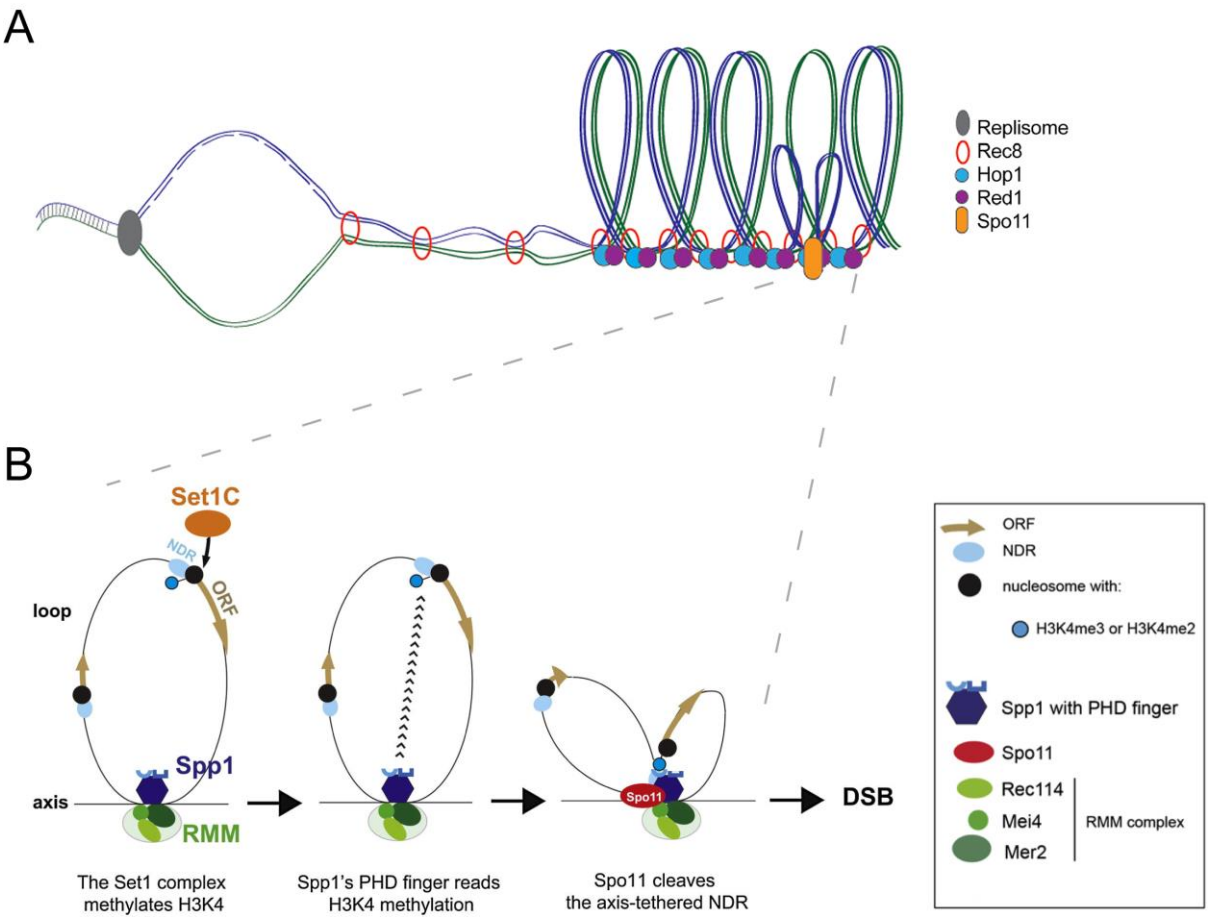


Figure 3. DSB formation in the context of a meiotic chromosome axis. Adapted from Lawrence et al. (2012) and Sommermeyer et al. (2012).

(A) During meiotic DNA replication, REC8-cohesin encloses the replicated chromatids and maintains sister cohesion. Rec8-cohesin associates with the budding yeast axis proteins Red1 and Hop1, orthologs of Arabidopsis ASY1 and ASY3, and forms the loop-axis structure. Spo11 then forms DSBs at the base of chromatin loops. (B) A schematic representation of mouse DSB formation according to the tethered loop/axis hypothesis. DSBs form in the nucleosome depleted regions (NDRs) upstream of genes. Spp1 is located on the chromosome axis, where its PHD domain interacts with H3K4me3, deposited by the Set1C methyltransferase. This association promotes DSB formation by Spo11. However, it is important to note that there is no direct evidence for the steps in the tethered loop/axis model being temporally ordered and/or dynamic.

1.1.7 Class I crossovers

After DNA synthesis, most joint molecule intermediates are thought to progress to non-crossovers via disassociation of the D-loop, without second-end capture (fig. 2G) (McMahill et al., 2007). However, a subset progress down either the Class I or Class II crossover pathways, which diverge from a common strand invasion intermediate (the D-loop) (fig. 2E) (Copenhaver et al., 2002; Bugreev et al., 2011). Class I crossovers display the phenomenon of crossover interference, whereby crossovers are more evenly spaced than expected by chance (Copenhaver et al., 2002). In addition, Class I crossovers depend on a conserved group of proteins, termed ZMM (Zip1, Zip2, Zip3, Zip4, Mer3, Msh4 and Msh5), which were first identified in budding yeast (Börner et al., 2004). Arabidopsis has functional orthologues of ZMM proteins that have been shown to play broadly conserved roles (Mercier et al., 2015).

For example, the synaptonemal complex protein Zip1 acts in budding yeast both locally to promote ZMM function and globally to support formation of the synaptonemal complex (Sym et al., 1993; Chen et al., 2015; Voelkel-Meiman et al., 2015). Recent work has implicated the ZZS (Zip2–Zip4–Spo16) complex in marshalling recombination intermediates towards crossovers in the context of a dynamic chromosome axis (De Muyt et al., 2018). Crucial to this process is Zip4, a large TPR repeat protein that functions as a direct physical bridge between the chromosome axis and crossover formation (De Muyt et al., 2018). For example, Zip4 is thought to mediate interactions between Zip2-Spo16, Zip3, Msh4, Msh5 and the axis protein Red1 (Arora & Corbett, 2018; De Muyt et al., 2018). An ortholog of Zip4 also exists in plants that is required for Class I crossovers (Chelysheva et al., 2007). Zip2-Spo16 resembles the structure-specific XPF–ERCC1 nuclease, whilst lacking enzymatic activity, and has been shown to bind branched DNA structures reminiscent of recombination intermediates (De Muyt et al., 2018). The Arabidopsis SHOC1 protein is an ortholog of budding yeast Zip2, and

interacts with the plant specific ZMM protein PTD, which also promotes Class I crossovers (Wijeratne et al., 2006; Macaisne et al., 2008; Macaisne et al., 2011). Thus, ZMM proteins coordinate crossover formation in the context of a dynamic chromosome axis.

In budding yeast, the DNA helicase Mer3 regulates the extension of nascent D-loops by DNA synthesis and stabilises JMs, as demonstrated with physical and genetic assays (Nakagawa & Kolodner, 2002; Mazina et al., 2004; Börner et al., 2004; Duroc et al., 2017), and an ortholog in Arabidopsis (*MER3*) also functions in the ZMM pathway (Chen et al., 2005; Mercier et al., 2005). Arabidopsis also has a functional ortholog of the yeast *ZIP3* E3 ligase gene, called *HEI10*, thought to regulate the distribution of crossovers through its role in mediating the turnover of meiotic recombination proteins (Agarwal & Roeder, 2000; Chelysheva et al., 2012; Reynolds et al., 2013; Qiao et al., 2014; Ziolkowski et al., 2017). With the exception of *mer3*, all Arabidopsis ZMM mutations reduce crossovers to ~15% of the wild type frequency, whereas crossovers in *mer3* remain higher at ~25% (Mercier et al., 2015). Hence, the majority of meiotic crossovers in Arabidopsis form via the Class I crossover pathway.

In Arabidopsis, Class I crossovers are primarily resolved by the endonuclease MutL Homolog 1 (MLH1)-MLH3 (MutLγ) heterodimer, as crossover frequency was reduced by 50% in Arabidopsis *mlh1* or *mlh3* mutants, whilst not reducing further when combined with additional *zmm* mutants (fig. 2F) (Jackson et al., 2006; Dion et al., 2007; Chelysheva et al., 2012). Immunostaining of Arabidopsis male meiocytes revealed that MLH1 and MLH3 are first detected in zygotene, and gradually increase in foci number to a mean of 9.4 per cell (Jackson et al., 2006; Lambing et al., 2015). In budding yeast, although Mlh1-Mlh3 does not have structure specific endonuclease activity *in vitro*, its polymerisation on DNA substrates was shown to induce nicks (Manhart et al., 2017), suggesting a potential mechanism for crossover resolution *in vivo*. Given the highly conserved structure of this heterodimer, a similar mechanism may function in resolving crossovers in Arabidopsis.

1.1.8 Class II crossovers

In Arabidopsis, the mechanism responsible for the residual 15% of ZMM-independent crossovers is poorly understood, and this pathway does not display the characteristic pattern of crossover interference observed for Class I crossovers (fig. 2J) (Copenhaver et al., 2002; Serra et al., 2018a). However, Class II crossovers are likely produced by multiple, overlapping, DNA repair processes, and crossovers occurring via this pathway are strongly suppressed by anti-crossover factors in wild type (Crismani et al., 2012; Fernandes et al., 2018a; Serra et al., 2018a).The only known protein specifically promoting crossovers in this pathway is MUS81 (Berchowitz et al., 2007; Higgins et al., 2008a; Crismani et al., 2012). Loss of MUS81 function causes a 10% reduction in crossover frequency, and the residual crossovers in a *zmm mus81*

double mutant are reduced by a further 30%, indicating other unknown mechanisms of Class II crossover resolution (Berchowitz et al., 2007; Higgins et al., 2008a).

1.1.9 The genetics of MutSγ

The ZMM proteins MSH4 and MSH5 are members of the MutS-homolog (*MSH*) gene family, and form a meiosis specific heterodimer termed MutSγ (Ross-Macdonald & Roeder, 1994). Unlike other members of the *MSH* family, studies in budding yeast indicated that MutSγ plays no role in mismatch repair (MMR), the process through which DNA mispairs (such as C:T or A:G) or small indels are identified and corrected (Ross-Macdonald & Roeder, 1994; Hollingsworth et al., 1995; Novak et al., 2001; Kunkel & Erie, 2015). Instead, MutSγ functions in promoting meiotic crossovers, via the Class I DNA repair pathway (Ross-Macdonald & Roeder, 1994; Hollingsworth et al., 1995; Novak et al., 2001). In budding yeast, the *msh4* mutation causes a loss of spore viability and defective chromosome segregation, indicative of failure in crossover formation (Ross-Macdonald & Roeder, 1994; Hollingsworth et al., 1995; Novak et al., 2001). In mouse, MutSγ foci appear during leptotene, peak in number during zygotene, and then decrease through to mid-pachytene (Kneitz et al., 2000). Analysis of RAD51 co-localisation indicated that MutSγ acts downstream of strand invasion, but is required for homologue pairing and synapsis (Kneitz et al., 2000).

In Arabidopsis, MSH4 forms numerous (80-100) axis-associated foci at mid-leptotene (Higgins et al., 2004). MSH4 foci gradually decline in number, with very few remaining by early-pachytene, and none remaining at late-pachytene (Higgins et al., 2004). MSH4 also co-localises with RAD51 in Arabidopsis, and there is a short window where RAD51 precedes MSH4 loading (Higgins et al., 2004). As expected, the *msh5* phenotype is almost identical to that of *msh4* in Arabidopsis and localisation of MSH5 is dependent on *MSH4* (Higgins et al., 2008b; Lu et al., 2008). Moreover, a subset of MSH5 foci also transiently co-localise with the Class I crossover marker MLH1 (Higgins et al., 2008b; Lu et al., 2008), suggesting that MutSγ promotes an association between crossover intermediates and crossover resolution factors. MutSγ also plays a conserved function in rice, where MSH4 and MSH5 localise as foci on meiotic chromosomes, and loss of function mutations cause reduced fertility (Luo et al., 2012, Zhang et al., 2014). These data are consistent with a model where MutSγ is targeted to recombination intermediates to promote crossover formation, and indicates that MutSγ plays a conserved role in plants.

Interestingly, the effect of losing MutSγ on synapsis is variable between species. For example, mutating the *MSH4* orthologue in *C. elegans* causes a near total loss of crossovers, but has no effect on synapsis (Zalevsky et al., 1999). This is consistent with earlier observations that synapsis is independent of recombination in *C. elegans* (Zetka & Rose, 1995; Dernburg et al.,

1998). The *msh4* budding yeast phenotype appears similar to Arabidopsis, where synapsis is mostly completed, but with local defects and unpaired regions (Ross-Macdonald & Roeder, 1994; Hollingsworth et al., 1995; Novak et al., 2001; Higgins et al., 2004). In contrast, *msh4* mutant mice show severe defects in synapsis (Kneitz et al., 2000). In mouse, approximately ~70% of pachytene *msh4* spermatocytes showed limited interhomolog pairing, and closer examination revealed that these interactions were between non-homologous chromosomes (Kneitz et al., 2000). Furthermore, mouse *msh4* chromosomes were not fully condensed, in comparison to the paired wild type chromosomes, which may be an indirect consequence of *msh4* impeding crossover maturation (Kneitz et al., 2000). Attempting to make sense of this variation, Higgins et al. (2004) suggested that differing synapsis phenotypes are likely due to variation in the mode of synaptonemal complex formation, rather than differences in the action of the highly conserved MSH4-MSH5 heterodimer.

1.1.10 The structure and function of MutSγ

MutSγ shares four of five domains (II-V) found in the MSH family proteins, including the structural domain containing the ATPase (fig. 4A,B) (Rakshambikai et al., 2013). However, both MSH4 and MSH5 lack the N-terminal domain I, thought to interact with domain IV to achieve mismatch recognition (Ross-Macdonald & Roeder, 1994; Hollingsworth et al., 1995). A predicted structure of the budding yeast Msh4-Msh5 heterodimer, based on homology modelling of the human MMR MSH2-MSH6 heterodimer (MutSα), revealed that MutSγ has a cavity volume of 16,676 $Å^3$, providing sufficient space for binding a dHJ (fig. 4B) (Rakshambikai et al., 2013). Thus, structural modelling of MutSγ is consistent with its proposed role in binding to crossover intermediates and promoting their resolution as crossovers.

In contrast to MutSα, purified human MutSγ specifically binds to Holliday junctions (HJs) *in vitro* (Snowden et al., 2004). Human MutSγ showed no affinity for any other DNA substrate tested, including dsDNA, mismatched dsDNA, and Y-junctions (Snowden et al., 2004). In addition, a recent study showed budding yeast MutSγ to have moderate affinity for 3'-overhangs, ssDNA forks, and displacement loops (Lahiri et al., 2018). A DNase I footprint assay performed on human MutSγ-bound HJ complexes revealed a 25 bp footprint over the core of the HJ, embracing two DNA duplexes (Snowden et al., 2004). Notably, this area of protection was consistent with earlier observations of bacterial MMR MutS homodimers (Su & Modrich, 1986) and MutSα protection assays (Gradia et al., 1997), and implies a conserved mode of DNA interaction within the *MSH* gene family. The presence of HJs stimulated MutSγ's ATPase activity and ATP binding induced the release of MutSγ as a 'sliding clamp', which exposed the core region to DNase I digestion (Snowden et al., 2004). These observations

suggest that MutSγ may promote crossover formation by physically stabilising recombination intermediates and/or protecting them from anti-crossover factors (fig. 4C).

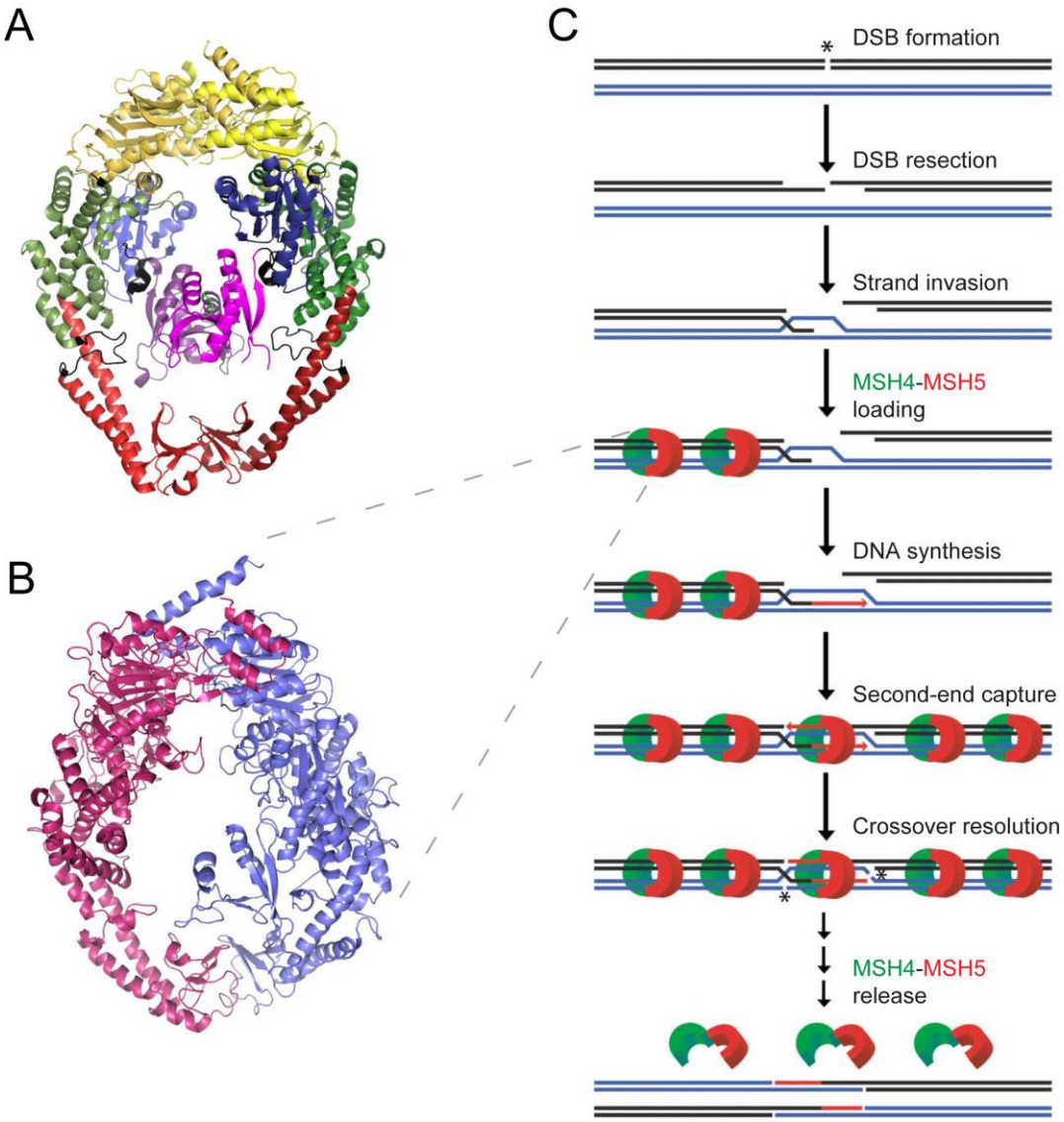


Figure 4. Structure of the MutS homodimer compared with the MutSγ heterodimer. Adapted from Rakshambikai et al. (2013).

(A) The crystal structure of *E. coli* MutS homodimer, including domain I (magenta), domain II (blue), domain III (green), domain IV (red), domain V (yellow) and connecting regions (black). (B) Model of the budding yeast MutSγ heterodimer. The binding of Msh4 (purple) and Msh5 (magenta) creates an enlarged central cavity, due to the absence of domain I. (C) Schematic representation of MutSγ loading onto strand invasion intermediates, following DSB formation and resection. MutSγ iteratively loads onto the intermediate and is released as a sliding clamp, enclosing two DNA duplexes. MutSγ is then released during crossover maturation and resolution.

Despite these findings, the *in vivo* DNA substrate for MutSγ binding remains unclear. The *in vitro* observations also raise the question of how clamping around two DNA duplexes can be achieved, and how MutSγ is loaded onto recombination intermediates. There also appears a difference in the conclusions reached from biochemical and cytological analyses (Higgins et al., 2004, 2008b; Luo et al., 2012, Zhang et al., 2014), with the latter suggesting that MutSγ is associated with recombinases and early recombination substrates, such as strand invasion intermediates. In contrast, *in vitro* binding assays show that MutSγ has a specific affinity for later dHJ substrates (Snowden et al., 2004; Lahiri et al., 2018). However, it is also possible that MutSγ is first recruited to recombination intermediates via its interactions with other proteins, and only later interacts physically with DNA.

1.1.11 Interactions with MutSγ

In support of a model where MutSγ is directly recruited by early recombination proteins, yeast two-hybrid and co-immunoprecipitation experiments revealed that human MSH4 physically interacts with RAD51 and DMC1 (Neyton et al., 2004), suggesting a role for MutSγ in the early stages of meiotic recombination. Furthermore, an *in vitro* study using GST-pull downs found evidence that human MSH4 interacts with recombinases RAD51D and XRCC3, and homologous recombination mediator protein RAD54 (Snowden et al., 2008). Unlike previous work, this study found no evidence for an interaction between MSH4 or MSH5 with RAD51, raising concerns about the reliability of *in vitro* data and its extrapolation to an *in vivo* context (Snowden et al., 2008). For instance, whilst *in vitro* experiments are performed in simplified and artificial reaction conditions, meiotic proteins interact in a highly complex chromatin environment *in vivo*. However, taken together these investigations suggest a role for MutSγ in mediating, and potentially physically stabilising, early meiotic recombination events.

In addition to its role in early recombination, MutSγ recruits and activates dHJ resolving factors, such as the Mlh1-Mlh3 endonuclease, both directly and indirectly (Santucci-Darmanin et al., 2000; Nishant et al., 2008; Zakharyevich et al., 2012; Manhart et al., 2017). Co-immunoprecipitation assays revealed that human MSH4 interacts directly with MLH1 *in vitro*, and the N-terminal region of MSH4 was responsible for this interaction (Santucci-Darmanin et al., 2000). Human MSH4 was also shown to interact with MLH3, and MLH3 co-immunoprecipitated with MSH4 from mouse meiotic cell extracts (Santucci-Darmanin et al., 2000). This is consistent with cytological data in Arabidopsis, where MLH1 foci co-localise with MutSγ (Higgins et al., 2008b), and with a model whereby MutSγ recruits MutLγ to promote crossover formation (Manhart et al., 2017).

Genetic and molecular assays indicate that MutSγ antagonises the activity of the budding yeast anti-crossover helicase Sgs1 (an ortholog of mammalian *BLM* and Arabidopsis *RECQ4A*

and *RECQ4B*) (Jessop et al., 2006). Although deleting *Sgs1* causes a relatively modest ~1.6-fold increase in crossover frequency in the *URA3-ARG4* interval, there is a greater 2-8-fold increase when *sgs1* is combined with *msh4*, relative to the *msh4* single mutant (Jessop et al., 2006). Furthermore, deleting Sgs1 in the *msh4* background also restores juxtaposition of the axes during synapsis (Jessop et al., 2006). Therefore, Msh4 antagonises the anti-crossover activity of Sgs1. Consistent with data in budding yeast, deleting the Drosophila BLM helicase restores an almost wild type crossover frequency in a *rec* mutant background (the *Drosophila* functional ortholog of MutSγ) (Kohl et al., 2012). This indicates a conserved role for MutSγ in antagonising anti-crossover helicases during meiosis.

1.1.12 Regulating the distribution of MutSγ

Several factors regulate MutSγ localisation and turnover on meiotic chromosomes (Reynolds et al., 2013; Qiao et al., 2014). The budding yeast protein Zip3 is a SUMO E3 ligase that promotes localization of ZMM proteins to recombination intermediates and facilitates synapsis (Agarwal & Roeder, 2000; Cheng et al., 2006; Macqueen & Roeder, 2009; Shinohara et al., 2008). HEI10, the Arabidopsis and rice ortholog of Zip3, is retained at a limited number of sites that correspond to Class I crossovers during the pachytene to diakinesis transition, where it co-localises with MLH1 (Chelysheva et al., 2012; Wang et al., 2012). The N-terminal RING-finger domain in budding yeast Zip3 is thought to catalyse the conjugation of ubiquitin-like molecules, such as Small Ubiquitin-Like Modifier (SUMO), to other proteins (Perry et al., 2005; Cheng et al., 2006). These findings implicate post-translational modifications in regulating crossover factors, potentially via the targeting of MutSγ for protein degradation.

Indeed, the mouse SUMO ligase RNF212 is also a member of the Zip3 class of E3 ligases, and its localisation to meiotic chromosomes is essential to stabilise a subset of MutSγ foci (Reynolds et al., 2013). Furthermore, the mouse ubiquitin ligase HEI10 is essential for crossover/non-crossover differentiation, and HEI10 foci occur at a subset of designated crossover sites at the pachytene stage in wild type (Qiao et al., 2014). In the absence of HEI10, RNF212 foci hyper-accumulate and block the progression of recombination intermediates (Qiao et al., 2014). Thus, RNF212 and HEI10 antagonise each other's function, resulting in robust crossover maturation, which implicates MutSγ as a key regulatory node in crossover/non-crossover discrimination. Both mouse *RNF212* and Arabidopsis *HEI10* show haploinsufficiency, and transformation of additional *HEI10* copies was sufficient to more than double crossovers in the chromosome arms in Arabidopsis (Reynolds et al., 2013; Ziolkowski et al., 2017).

The proteasome is a multi-subunit protease that degrades both nuclear and cytoplasmic proteins, and specifically targets proteins tagged with ubiquitin, but not those tagged with

SUMO, for degradation (Rao et al., 2017). Meiotic recombination stalls when the proteasome is deactivated in mouse spermatocytes, reminiscent of the *hei10* phenotype (Qiao et al., 2014; Rao et al., 2017). Furthermore, in the *hei10* mutant MutSγ persists at sites that would normally form non-crossovers, consistent with MutSγ normally being targeted for proteolysis at these sites (Rao et al., 2017). In support of a major role for proteolytic degradation during meiotic recombination, both the proteolytic core of the proteasome and its regulatory particles are recruited to budding yeast chromosomes by Zip3 and Zip1 (Ahuja et al., 2017).

Interestingly, in budding yeast MutSγ is initially unable to promote crossover formation, due to an N-terminal degron on Msh4 that renders it unstable (He et al., 2018). Stabilisation of MutSγ requires Cdc7, which directly phosphorylates, and thereby neutralises, the Msh4 degron (He et al., 2018). Targeted mutagenesis of putative phosphorylation residues causes a loss of crossover interference and reduced crossover frequency, phenocopying the *msh4* deletion, and the *msh4* phosphorylation mutant could be rescued through proteasome inhibition by MG132 (He et al., 2018). This discovery substantiates previous studies implicating the ubiquitin-proteasome system in crossover formation, via the targeted (de)stabilisation of MutSγ (Reynolds et al., 2013; Qiao et al., 2014; Ahuja et al., 2017; Rao et al., 2017; He et al., 2018).

1.1.13 MutSγ and crossover interference

Residual crossovers in the budding yeast *msh4* null mutant display significantly reduced crossover interference (Novak et al., 2001). This raises the question of whether MSH4-MSH5 are themselves responsible for interference, or are simply required for the Class I crossover formation (Novak et al., 2001; Mancera et al., 2008). It was argued that local depletion of MutSγ, through iterative loading of the heterodimer, causes interference via a ‘first come, first served’ mechanism (Kunz & Schär, 2004). Such a model is entirely consistent with the lack of crossover interference in fission yeast, which lacks MutSγ (Munz, 1994). In support of this, interference was observed between mouse MSH4 foci at late zygotene, as well as RPA foci, in addition to interference between MLH1 foci (de Boer et al., 2006). Global microarray analysis of the budding yeast recombination landscape showed that deleting *msh4* had no effect on non-crossover numbers (Chen et al., 2008). These results suggest that MutSγ marks interfering recombination intermediates, of which a subset are destined for crossover formation. However, budding yeast Zip2 foci also display an interference pattern, despite loading in advance of other ZMM factors, including MutSγ, suggesting that crossover interference is operating before MutSγ localises to chromosomes (Fung et al., 2004). These data argue against MutSγ being the primary determinant of crossover interference. Hence, the role of MutSγ in mediating crossover interference appears to vary between species.

1.1.14 Meiotic anti-crossover factors

In Arabidopsis, crossovers are outnumbered by DSBs, implying that many are resolved as non-crossovers, where recombination intermediates are resolved without reciprocal exchange of flanking chromosomes, or via inter-sister repair (Sun et al., 2012; Yang et al., 2012; Drouaud et al., 2013; Wijnker et al., 2013). Non-crossovers are thought to form via the synthesis dependent strand annealing (SDSA) pathway, which is analogous to the mitotic SDSA pathway that functions to repair mitotic DSBs occurring due to stalled replication forks or exogenous DNA damage agents (Heyer et al., 2010). In SDSA, DNA synthesis first extends a D-loop intermediate to generate additional sequence, with complementarity to the other DSB end, and the D-loop is then disassembled (McMahill et al., 2007). This leads to annealing of the nascent strand to a region of homology exposed during resection of the second end, resulting in a nonexchange event (McMahill et al., 2007).

In Arabidopsis, several partially redundant anti-crossover pathways act to promote non-crossover formation (Mercier et al., 2015). The RECQ4 complex is orthologous to the highly conserved human BTR complex (BLM-TOP3-RMI1-RMI2) and the yeast Sgs1-Top3-Rmi1 complex, and disrupts D-loops and dHJs *in vitro* (Fasching et al., 2015). Arabidopsis has two BLM orthologues, RECQA and RECQB, with redundant functions (Hartung et al., 2007; Mannuss et al., 2010; Schröpfer et al., 2014). Two recent studies demonstrated a massive increase in crossover frequency in the *recq4a recq4b* double mutant (Fernandes et al., 2018a; Serra et al., 2018a), although this increase was in contrast to the more modest meiotic crossover increases in budding yeast *sgs1* mutants (Jessop et al., 2006). The meiotic function(s) of TOP3α and RMI1 roles are harder to disentangle, due their roles in both crossover regulation and resolving chromosome interlocks (Séguéla-Arnaud et al., 2017; Wang & Copenhaver, 2018).

The FANCM DNA helicase and its cofactors, MHF1 and MHF2, form a second conserved anti-crossover complex (Crismani et al., 2012; Girard et al., 2014). FANCM activity promotes dHJ branch migration *in vitro* (Gari et al., 2008), and was shown to regulate the non-crossover pathway in fission yeast (Lorenz et al., 2012). A genetic screen in Arabidopsis that aimed to identify mutations rescuing the infertility phenotype of the *zip4* mutant, caused by reduced Class I crossovers, found mutations in *FANCM* and its two cofactors (Crismani et al., 2012; Girard et al., 2014). The *fancm* mutant had a 3-fold increase in crossovers, and equivalent increases were observed in the *mhf1* and *mhf2* backgrounds (Crismani et al., 2012; Girard et al., 2014). These increase were exclusively via the Class II crossover pathway, as crossover increases in *fancm* showed no crossover interference and did not require *ZIP4* (Crismani et al., 2012).

The final group of anticrossover factors are the AAA-ATPase proteins FIGL1 and its partner FLIP (Girard et al., 2015; Fernandes et al., 2018b). In Arabidopsis, knockouts of either gene lead to significant increases in meiotic crossover frequency, with *figl1* and *flip* mutations causing a 70% and 30% increase, respectively (Girard et al., 2015; Fernandes et al., 2018b). FIGL1 was observed to regulate DMC1 and RAD51 localisation, with *figl1* mutants having a two-fold increase of the number of RAD51 foci, whilst the number of DMC1 foci was unchanged (Girard et al., 2015; Fernandes et al., 2018b). However, DMC1 foci persisted into pachytene in *figl1* cells, but not in the wild type (Girard et al., 2015). Furthermore, FIGL1 was shown to directly interact with RAD51 and DMC1 in a yeast two hybrid assay (Fernandes et al., 2018b). Because the *figl1 mus81* double mutation caused entangled meiotic chromosomes and sterility, the authors proposed that FIGL1 prevents the formation of aberrant recombination intermediates, whose resolution by MUS81 otherwise leads to additional crossovers (Fernandes et al., 2018b). Hence, FIGL1-FLIP likely acts to coordinate, as well as limit, interactions between homologous chromosomes during the early stages of meiotic recombination (Girard et al., 2015).

1.1.15 The regulation of crossover landscapes

Crossover rates are not homogenous across the genome and chromosomes have alternating regions of high and low crossover frequency (Drouaud et al., 2006; Giraut et al., 2011; Wijnker et al., 2013; Serra et al., 2018a). For instance, in budding yeast, humans, wheat, and Arabidopsis, more than 80% of crossovers occur in less than a quarter of the genome (Myers et al., 2005; Chen et al., 2008; Mancera et al., 2008; Choi et al., 2013; International Wheat Genome Sequencing Consortium, 2014). However, despite striking variation in crossover frequency across the genome, almost all eukaryotes display crossover homeostasis and crossover assurance. The former is the property of total crossover number being maintained despite changes to earlier recombination, such as altered DSB levels (Martini et al 2006; Cole et al., 2012; Xue et al., 2018), whilst crossover assurance is the property of all chromosomes being guaranteed at least one crossover per meiosis (Shinohara et al., 2008). These properties indicate that meiotic recombination is regulated at multiple levels.

DNA methylation has been shown to impact crossover distributions in Arabidopsis (Melamed-Bessudo & Levy, 2012; Mirouze et al., 2012; Yelina et al., 2012; Yelina et al., 2015; Underwood et al., 2018). In Arabidopsis, DNA methylation occurs in three sequence contexts, CG, CHG and CHH, and is elevated in proximity to the centromeres (Stroud et al., 2013, 2014). The Arabidopsis *met1* or *ddm1* methyltransferase mutations, which predominantly reduce CG context DNA methylation, cause crossover redistribution away from the centromere, despite total crossover events remaining unchanged (Mirouze et al., 2012; Yelina et al., 2012;

Melamed-Bessudo & Levy, 2012). Heterochromatin is generally refractory to meiotic recombination (Ellermeier et al., 2010; Choi et al., 2018; Underwood et al., 2018), and RNA directed DNA methylation is sufficient to suppress crossover hotspot activity at the fine-scale (Yelina et al., 2015), by recruiting dense DNA methylation, H3K9me2 and high nucleosome density. As a consequence, large portions of the Arabidopsis genome are silent for both recombination initiation (Choi et al., 2018; Underwood et al., 2018), and crossover formation (Copenhaver et al., 1998; Copenhaver et al., 1999; Giraut et al., 2011; Serra et al., 2018a).

During meiosis, cohesin is enriched around the centromere and adjacent pericentromeric heterochromatin (Bernard et al., 2001; Blat et al., 2002; Klein et al., 1999; Mizuguchi et al., 2014; Sun et al., 2015; Tanaka et al., 1999; Watanabe & Nurse, 1999), where it is known to suppress crossovers. For example, in fission yeast the meiosis-specific Rec8-Psc3 complex suppresses recombination in the pericentromeres (Nambiar & Smith, 2018). Likewise, the Arabidopsis REC8 protein (a meiosis-specific α-kleisin cohesin subunit) is enriched in proximity to the centromere, where crossovers and DSBs are suppressed (Lambing et al., 2019). Consistent with this suppressive role, reducing H3K9me2 and non-CG DNA methylation in a *kyp suvh5 suvh6* triple mutant caused a redistribution of REC8 and a subsequent increase in meiotic recombination within repeat sequences (Underwood et al., 2018; Lambing et al., 2019). Hence, cohesin and heterochromatin may act to synergistically repress meiotic crossovers around the centromere, in order to limit recombination between repetitive centromere-associated DNA and to ensure sister chromatid cohesion remains intact.

Positive correlations occur between crossovers and the euchromatin mark H3K4me3, histone variant H2A.Z, and DNA hypomethylation in plants (Cao et al., 2011; Choi et al., 2013; Long et al., 2013). In Arabidopsis, the +1 nucleosome is associated with elevated historical crossover frequency, and also has high levels of H3K4me3 and H2A.Z (Choi et al., 2013). Indeed, in the *arp6* mutant, which is defective in H2A.Z deposition, crossover frequency was reduced at crossover hotspots and genome-wide (Choi et al., 2013). Genome-wide DSB maps in budding yeast, maize, and Arabidopsis have revealed that hotspots are concentrated in the nucleosome-depleted regions of genes (Pan et al., 2011; Lam & Keeney, 2014; He et al., 2017; Choi et al., 2018). Furthermore, several euchromatic histone modifications are associated with DSB hotspots: a) H3K4me3 in yeast, mice, humans, and Arabidopsis, b) histone H3 lysine 9 acetylation (H3K9ac) in fission yeast, and c) H3 lysine 5 acetylation (H3K5ac) in *C. elegans* (Mézard et al., 2015). Consistent with these findings, elevated crossover frequency is found in the nucleosome depleted regions of gene promoters, in plants, mammals and birds (Auton et al., 2013; Choi et al., 2013; Wijnker et al., 2013; Singhal et al., 2015; Demirci et al., 2017). Together, these observations demonstrate that chromatin is an important determinant of meiotic recombination landscapes in Arabidopsis and other eukaryotes.

In addition to epigenetic factors, DNA sequence motifs associate with both DSBs and crossovers (Choi et al., 2013; Wijnker et al., 2013; Shilo et al., 2015). For example, crossovers occur at predominantly AT-rich sequences in plants, likely due to their association with nucleosome depletion (Choi et al., 2013). As AT-rich sequences also correlate with elevated SPO11-1-dependent meiotic DSBs, this likely explains this positive association (Choi et al., 2018). Two further sequence motifs, a CTT-repeat and CNN-repeat, associate with crossover hotspots, such as *RAC1* (Horton et al., 2012; Wijnker et al., 2013; Shilo et al., 2015; Choi et al., 2016). These trinucleotide repeats associate with both promoters and the +1 nucleosome of gene bodies, and correlate with crossover frequencies derived from both direct experimental and historical approaches (Cao et al., 2011; Choi et al., 2013; Long et al., 2013; Choi et al., 2016; Serra et al., 2018b). As these repeats associate with underlying chromatin features, it remains an open question whether they directly influence crossover formation or are simply correlated with crossover frequency.

At the broad-scale, budding yeast DSB and crossover densities positively correlate (Mancera et al., 2008; Pan et al., 2011). In Arabidopsis, the correlation between DSBs and crossovers is positive at the chromosome-scale, but less robust at the fine-scale (Choi et al., 2018; Serra et al., 2018b). However, this may be a consequence of the necessity of studying crossover distributions in a hybrid, where interhomolog DNA polymorphisms may impact DSB formation and repair, whereas SPO11-1-oligos have only been mapped in a homozygous background (Choi et al., 2018; Underwood et al., 2018). Interestingly, fine-scale analysis of crossover distributions at meiotic crossover hotspots indicates that crossovers localise to the 5'-end of gene bodies (Choi et al., 2013; Serra et al., 2018b). For example, at the Arabidopsis *RAC1* and *RPP13* crossover hotspots, SPO11-1-oligos were enriched in the nucleosome depleted promoter regions, whilst crossovers were enriched within the gene bodies (Serra et al., 2018b), as was previously reported (Choi et al., 2016). Strikingly, crossovers were anti-correlated with interhomolog polymorphism at the fine-scale, indicating that sequence divergence within recombination intermediates may inhibit crossover formation (Serra et al., 2018b). The role of interhomolog polymorphism in shaping the meiotic recombination landscape will be further explored in section 1.3.

In addition to the regulation of DSB formation by genome features, DSBs feedback on themselves to regulate DSB distributions (Cooper et al., 2016). In budding yeast, DSBs were shown to display distance-dependent interference, in *cis* (Garcia et al., 2015). DSB interference is the phenomenon whereby the formation of one DSB represses adjacent break formation (Cooper et al., 2016). This effect was observed over physical distances of ~70-100 kb, possibly corresponding to meiotic chromatin loops, and to rely on the conserved DNA damage response kinase Tel1 (Garcia et al., 2015). This mechanism is thought to suppress

the co-occurrence of multiple DSBs within the same genomic locus, which risks producing deletions and genomic instability (Cooper et al., 2016). Moreover, downstream steps in meiotic recombination feedback onto DSB formation (Thacker et al., 2014). For example, DSB density is elevated in the *zip1*, *zip3* and *msh5* backgrounds, indicating that meiotic recombination feeds back to suppress additional DSB formation (Thacker et al., 2014). Together, these observations support a hierarchical view of recombination initiation, where multiple levels of feedback act to both disperse recombination events and ensure obligate crossovers, ultimately supporting efficient genetic recombination whilst protecting genome stability.

1.2 Mismatch repair: a brief introduction

1.2.1 The generation of mismatched base pairs

DNA mismatch repair functions to correct errors arising from the misincorporation of bases during DNA replication and DNA damage, and therefore functions to maintain genome stability (Kunkel & Erie, 2015). Before polymerase errors are targeted by MMR, mismatched bases may stimulate the intrinsic proofreading activity of the polymerases themselves (Ozawa et al., 2008). Exonucleolytic proofreading is stimulated by the changed geometry of mismatched bases, which slows the polymerase's processivity (Johnson & Beese, 2004). However, several types of error manage to escape detection. Notably, the Hoogsteen base pair (8-oxo-G-dA) is stabilised by hydrogen bonds in the DNA major groove (Hoogsteen, 1963; Kunkel & Erie, 2015). As the geometry of this base pair is similar to the T-A base pair, three different eukaryotic polymerases incorporate dATP opposite 8-oxo-guanine, because they cannot distinguish between 8-oxo-guanine and a thymine base (Shibutani et al., 1991; Haracska et al., 2000; Sabouri et al., 2008). Therefore, despite this combination of factors acting to limit mismatch accumulation during replication, mismatches still occur *in vivo* and must be repaired in order to maintain genome stability.

1.2.2 MMR in bacteria

MMR proteins were first identified as 'mutation avoidance' factors, through screens for increased mutability in bacteria (for a historical perspective, see Modrich, 1987). Early genetics and biochemistry implicated ten proteins to function in bacterial MMR: MutS, MutL, MutH, DNA helicase II (MutU or UvrD), single-stranded binding protein (SSB), exonuclease I, exonuclease VII, RecJ exonuclease, DNA polymerase III, and DNA ligase (Modrich, 1991). Of these proteins, only MutS could interact with mismatches independent of any other factor, and its relative binding affinity for a particular mispair correlated with the efficiency of MMR (Modrich, 1991).

After mismatch recognition, MMR must discriminate between the template and nascent daughter strand, which by definition contains the replication error. In bacteria, MMR identifies

the nascent strand via its transiently undermethylated d(GATC) sequences (Wagner & Meselson, 1976; Pukkila et al., 1983), and repair is directed to the unmethylated strand (Herman & Modrich, 1981; Marinus et al., 1984; Urig et al., 2002). A single d(GATC) site is sufficient for discrimination, and distances of up to 1 kb can separate the d(GATC) site from the mismatch (Lahue et al., 1987; Bruni et al., 1988).

After mismatch recognition, the d(GATC) signal induces a mismatch dependent incision on the unmethylated strand (Modrich, 1991). The *MutL* gene product directly interacts with MutS, dependent on MutS hydrolysing ATP (Grilley et al., 1989; Mankovich et al., 1989). This interaction stimulates the latent endonuclease activity of the type II family endonuclease MutH, which incises the nascent strand at the d(GATC) site, 5′ to the G base (Welsh et al., 1987). The DNA terminus generated by MutH serves to direct MutL-dependent loading of DNA helicase II and SSB (Dao & Modrich, 1998). In a concerted reaction, these proteins unwind the DNA duplex and generate ssDNA (Dao & Modrich, 1998). This becomes the substrate for either 5′ or 3′ exonuclease digestion, as demonstrated by the presence of mismatch-dependent ssDNA 'excision' tracts in *Escherichia coli*, generated during MMR on a plasmid substrate in a polymerase suppressed system, which prevented re-synthesis of the excised regions (Su et al., 1989). These experiments showed excision tracts to run from the d(GATC) site to shortly past the mismatch (Su et al., 1989). After degradation, DNA Pol III resynthesises the DNA tract and DNA ligase seals the junction (Lahue et al., 1989).

1.2.3 MMR in eukaryotes

In spite of the increased complexity of eukaryotic DNA replication – such as multi-subunit DNA polymerases, multiple origins of replication, shorter and more frequent Okazaki fragments, and higher order DNA packing – eukaryotic MMR shows significant functional similarity to bacterial MMR (Kunkel & Erie, 2015). Unlike MMR initiation in bacteria, which is performed by homodimers of mismatch recognition protein MutS, eukaryotes have duplicated and diversified MutS activities (Romanova & Crouse, 2013). In mammals, two distinct MSH heterodimers are formed, each requiring MSH2 as a partner: MSH2-MSH6 (MutSα) and MSH2-MSH3 (MutSβ) (Modrich, 2006). In budding yeast, it was shown that single indel mutations in homopolymers are particularly prevalent in MMR mutants, through direct PCR-based measurements of microsatellite length (Sia et al., 1997; Lujan et al., 2014). In humans, 80-90% of cellular MSH2 forms a MutSα complex, which shows high specificity towards base-base mismatches and one or two nucleotide indels (Drummond et al., 1995; Palombo et al., 1996). MutSβ recognises single nucleotide indels with weak affinity and larger indel loops of 2-10 nucleotides with high affinity (Palombo et al., 1995; Genschel et al., 1998). Therefore, eukaryotes have subfunctionalised the mismatch recognition step of MMR.

MSH proteins belong to the ATP Binding Cassette (ABC) ATPase family, and their primary dimerization site is between their ATPase domain(s) (fig. 5A) (Groothuizen & Sixma, 2016). Mismatch recognition is conferred by the N-terminal mismatch binding domain of either MSH6 or MSH3, in MutSα and MutSβ, respectively (Groothuizen & Sixma, 2016). This domain is followed by the connector domain, which is important for MLH interactions, and the Core/Lever and Clamp domains, which are crucial for DNA interactions *per se* (Groothuizen & Sixma, 2016). This is followed by a helix-turn-helix (HTH) motif, which is required for dimer stability and also interacts with the ATPase domain of an opposing subunit (Groothuizen & Sixma, 2016). The final domain, notably absent from MSH6, is a tetramerisation domain (Groothuizen & Sixma, 2016). Evidence exists for transient tetramerisation for bacterial MutS proteins in solution, however this has not been demonstrated in eukaryotes (Groothuizen et al., 2013).

Obmolova et al. (2000) produced the first crystal structure of a bacterial MutS homodimer (fig. 5B,C). In the absence of DNA, the clamp domains showed disorder, indicating that conformational flexibility may be required to open and close the clamp domain when loading or unloading from DNA (Obmolova et al., 2000). In contrast, DNA binding induced a stabilised closed state. The later crystal structures of bacterial and eukaryotic mismatch-associated homodimers or heterodimers indicated a conserved mode of mismatch binding (fig. 5C-E) (Groothuizen et al., 2013). The binding of *E. coli* MutS and human MutSα cause a significant ~60° DNA kink, thought to result from the dimer 'sampling' the reduced helical stability caused by a mismatch (Obmolova et al., 2000; Groothuizen & Sixma, 2016).

The ATPase activity of MutS is required for MMR *in vivo*, and ATPase domain mutants retain their association with mismatches (Haber et al., 1988). Furthermore, these mutants were dominant over a wild type MutS in *E.coli*, and caused MMR efficiency to reduce by 12-90% (Modrich, 1991), suggesting that MutS remains associated with DNA in the absence of ATP hydrolysis, potentially blocking mismatch processing. Notably, DNA footprinting assays, in which DNA substrates are digested by DNase I to reveal the binding locations of proteins, indicated that MutS protected ~24-28 bp of DNA duplex (Biswas & Hsieh, 1997). However, the addition of both MutL and ATP increased the protected region to ~100 bp (Su & Modrich, 1986; Su et al., 1988). Together these observations suggest that MutS may be released as a sliding clamp surrounding DNA (hetero)duplexes, which is required for recruiting and stabilising MutL, as well as downstream MMR factors (fig. 5F) (Acharya et al., 2003). This 'sliding clamp' model was confirmed by single-molecule optical microscopy, where budding yeast Msh2-Msh6 was visualised sliding along homologous DNA via one-dimensional diffusion (Jiang et al., 2005; Gorman et al., 2007).

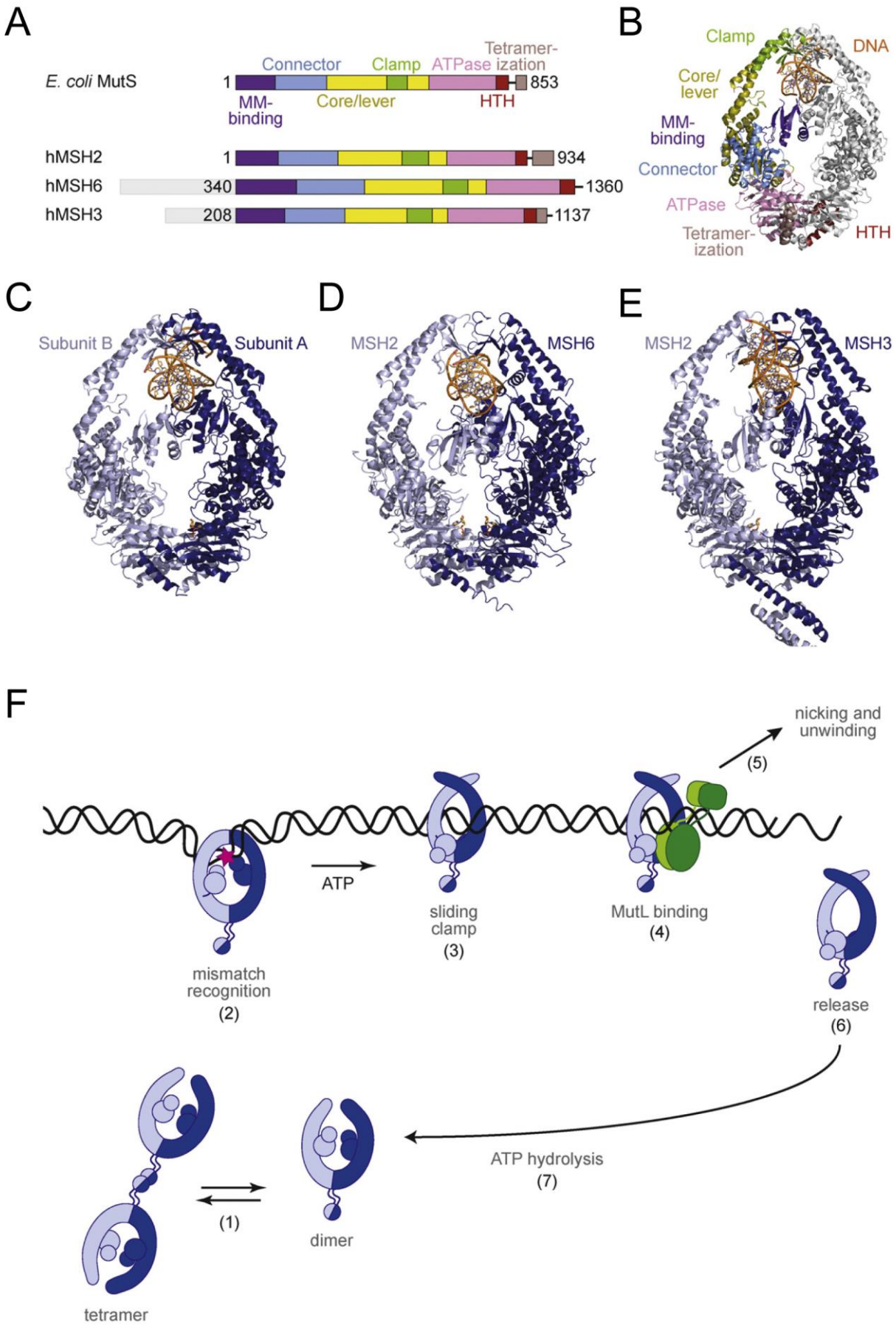

A
E. coli MutS
1
853
Connector
Clamp
ATPase
Tetramer-ization
MM-binding
Core/lever
HTH
hMSH2
1
934
hMSH6
340
1360
hMSH3
208
1137
B
Clamp
DNA
Core/lever
MM-binding
Connector
ATPase
Tetramer-ization
HTH
C
Subunit B
Subunit A
D
MSH2
MSH6
E
MSH2
MSH3
F
nicking and unwinding
(5)
ATP
sliding clamp
(3)
MutL binding
(4)
mismatch recognition
(2)
release
(6)
ATP hydrolysis
(7)
(1)
dimer
tetramer

Figure 5. MutS homolog domain organisation and structure. Adapted from Groothuizen & Sixma (2016).

(A) A comparison of *E. coli* and human MSH protein domains, showing the conserved domain organisation. The mismatch (MM) binding and helix-turn-helix (HTH) domains are indicated. (B) The crystal structure of *E. coli* MutS, with the domains coloured according to (A). The crystal structures are shown for G:T mismatch bound *E. coli* MutS (C) and human MutSα (D), and for 3 bp indel bound human MutSβ (E). (F) Schematic representation of MutS interactions with DNA. (1) MutS alternates between dimer and tetramer states in solution. (2) Mismatch (red star) recognition results in DNA kinking and retention at the mismatch site. (3) ATP binding induces the release of MutS as a sliding clamp. MutS interacts with MutL (4), and initiates MMR (5). (6) MutS is released from DNA, potentially dependent on ssDNA and/or DNA repair processes. (7) ATP is hydrolysed, returning MutS to its initial state (1).

In eukaryotes, MutL roles are played by three MLH heterodimers, each of which shares MLH1 as a binding partner: MLH1-PMS2 (MutLα), MLH1-PMS1 (MutLβ) and MLH1-MLH3 (MutLγ) (Kunkel & Erie, 2015). MutLα plays the major role in MMR, interacting with both MutSα and MutSβ (Li & Modrich, 1995). Although MutLβ or MutLγ have been shown to repair indels *in vitro* (Zhang et al., 2005; Jiricny, 2013), the *in vivo* effects of removing either MSH3, PMS1 or MLH3 have been more challenging to establish, given their reduced abundance, more specialised roles and possible redundancy (Kunkel & Erie, 2015).

Unlike bacteria, strand discrimination in eukaryotes does not rely on hemimethylation (Modrich & Lahue, 1996), but how the nascent strand is discriminated in eukaryotic MMR is still an area of debate (Kunkel & Erie, 2015). However, replication-associated MMR may not require an intermediary strand signal, given that replication in eukaryotes generates extensive DNA ends as a consequence of Okazaki fragments associated with lagging-strand replication (Modrich & Lahue, 1996). Consistent with this idea, an assay incorporating mutation reporter alleles in two orientations on opposite sides of a replication origin confirmed that MMR is more efficient on the lagging strand (Pavlov et al., 2003). Furthermore, this bias was lost in *msh2*, *msh6*, *mlh1* and *exo1* mutants, confirming that MMR activity was responsible for the bias (Pavlov et al., 2003). Interestingly, PCNA is known to be loaded asymmetrically at the replication fork, with a higher density on the lagging strand, potentially providing an additional mechanism to discriminate the nascent daughter strands (Kadyrova & Kadyrov, 2015). Indeed, the budding yeast Msh2-Msh6 heterodimer is associated with replication centres during S-phase, and MSH6-GFP foci reduce by 60% in a PCNA mutant (*Pol30-204*) previously identified as a mutator allele (Hombauer et al., 2011). Together, these observations highlight connections between DNA replication and MMR, and suggest that asymmetries in DNA replication forks may play a strand discrimination role in MMR.

Indeed, an intimate association between MMR and DNA replication in eukaryotes makes sense, when considering the ratio of protein subunits to replication structures (Kumar et al., 2011; Ghaemmaghami et al., 2013). An estimated ~1,300 Msh2 proteins are present in a budding yeast cell, whilst the 200 origins simultaneously active during S-phase give rise to 400 replication forks (Kumar et al., 2011; Ghaemmaghami et al., 2003). Hence, there is likely a need to associate MMR proteins with replication structures prior to the actual occurrence of biosynthetic errors, in order to rapidly detect and repair errors (Kunkel & Erie, 2015).

After strand discrimination, excision proceeds from the DNA nick (or end) and extends ~90-170 nucleotides past the mismatch (Fang & Modrich, 1993; Wang & Hays, 2004), as evidenced by radiolabelling DNA synthesis tracts after MMR correction (Thomas et al., 1991). Budding yeast and mammals have both Exo1-dependent and Exo1-independent MMR pathways (Kunkel & Erie, 2015). The former relies on the 5′ to 3′ hydrolytic activity of Exo1, which requires the functional integrity of the MutLα endonuclease metal-binding motif (Kadyrov et al., 2009). In the absence of Exo1, a second mechanism of mismatch processing requires a 5′ DNA end and strand-displacement synthesis by Pol δ or Pol ε (Kadyrov et al., 2009). A third Exo1-independent mechanism requires the 3′-5′ exonuclease activity of Pol δ or Pol ε (McCulloch et al., 2004). For example, even when stimulated with dNTPs, Pol δ or Pol ε could excise mismatches contained in primers, up to seven base pairs away from primer's 3′ end (McCulloch et al., 2004). After removing the error-containing nascent strand, gaps are filled by Pol δ, and sealed by DNA ligase (fig. 6) (Kunkel & Erie, 2015).

1.2.4 Plant MutS orthologs promote genome stability

Like other eukaryotes, Arabidopsis contains multiple *MSH* family genes, which group into conserved gene families (Culligan et al., 2000). Four MMR paralogues exist in Arabidopsis: *MSH2*, *MSH3*, *MSH6* and *MSH7* (Adé et al., 1999; Culligan & Hays, 2000). Consistent with other eukaryotic MSH-heterodimers, *in vitro* binding assays revealed that Arabidopsis MSH proteins have differing mismatch specificities: MutSα binds strongly to mismatches and single nucleotide indels, whilst MutSβ shows preference for larger indel loops (Wu et al., 2003). The plant specific MSH2-MSH7 heterodimer was found to bind a specific subset of mismatches *in vitro*, with highest affinity for G/G, G/A, A/A and particularly C/A mismatches (Wu et al., 2003).

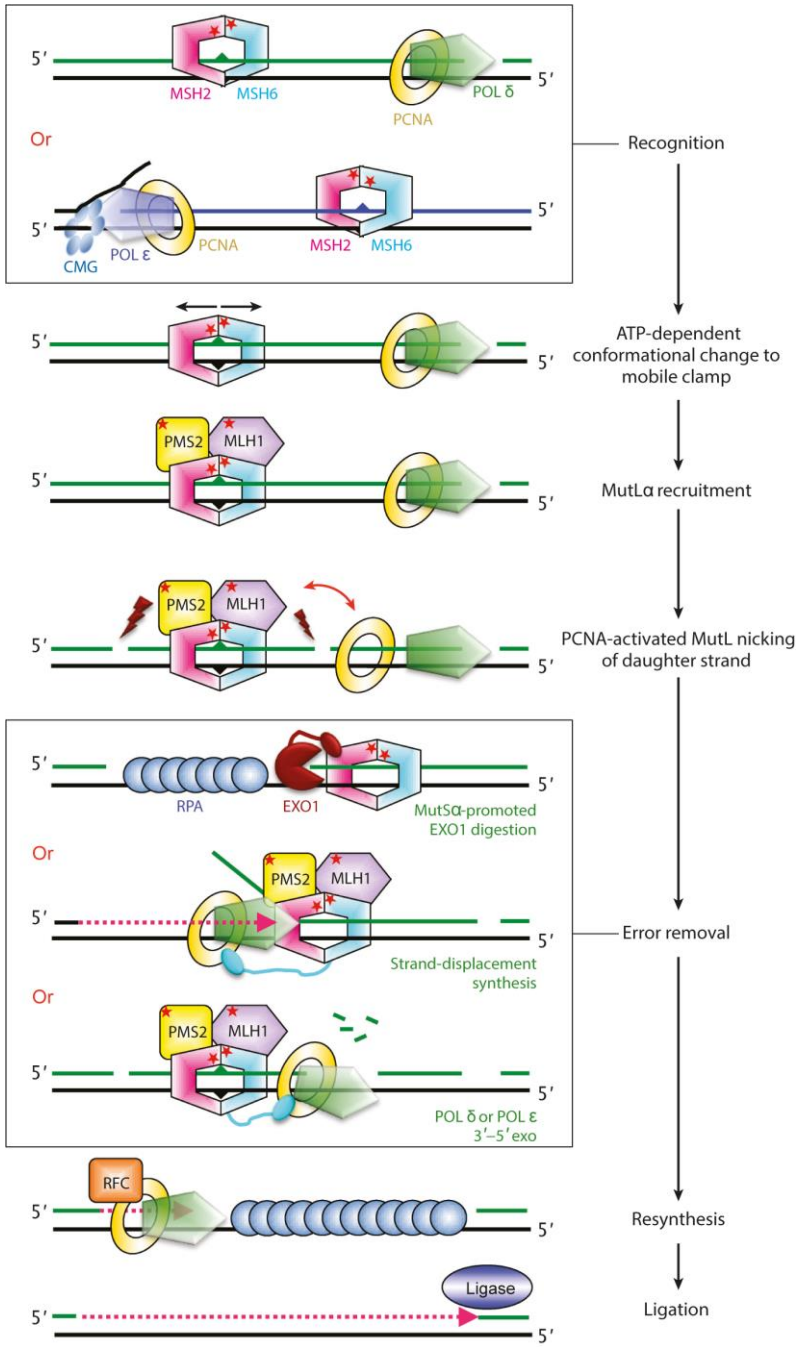


Figure 6. A summary model of replication-associated MMR in eukaryotes. Adapted from Kunkel & Erie (2015).

MMR initiates with MSH2-MSH6 (MutSα) binding to a mismatch resulting from polymerase misincorporation, on either the leading (bottom) or lagging (top) strand. Leading strand synthesis is performed by polymerase ε in conjunction with the CMG replicative helicase, and lagging strand synthesis is performed by polymerase δ. MutSα is then released as a sliding clamp, in an ATP-dependent manner, and recruits PMS2-MLH2 (MutLα). After PCNA induced nicking in the nascent strand, the error is removed via (i) EXO1 resection, with the ssDNA stabilised by RPA (ii) strand-displacement synthesis, or (iii) polymerase δ or polymerase ε 3′-5′ nuclease activity. The complementary strand is then re-synthesised, in association with replication factor C (RFC), and ligated.

In Arabidopsis, a *MSH2* knockout (*msh2-1*) causes endogenous microsatellite and repeat instability, when monitored using an out-of-frame *GUS* transgene reporter, where mutations due to repeat instability cause GUS positive sectors (Leonard et al., 2003). Furthermore, the propagation of *msh2-1* lineages results in severe developmental defects (Hoffman et al, 2004). For example, after five generations only 2 of 36 *msh2-1* lines retained a wild type phenotype (Hoffman et al, 2004). At the molecular level, the mutation rate increases ~100-fold in Arabidopsis *msh2-1* mutant accumulation lines (Watson et al., 2016; Belfield et al., 2018). Similar genomic instability phenotypes have been observed when disrupting *MSH* genes in a number of other plant species, including tomato (Sarma et al., 2018), tobacco (Van Marcke & Angenon, 2013) and moss (Trouiller et al., 2006). In conclusion, plant *MSH*-homologs show a high degree of sequence identity with their eukaryotic counterparts and display both a similar mismatch recognition spectrum and MMR-defective phenotypes.

1.2.5 MMR and the chromatin environment

Evidence has accumulated that MMR locally modifies the chromatin environment, in the context of both replication- and recombination-associated repair (Hauer & Gasser, 2017). Alignments of human MSH6 revealed this protein to share five domains with members of the MutS family, but that it also had a unique N-terminal region (Clark et al., 2007). These several hundred amino acids form a globular Pro-Trp-Trp-Pro (PWWP) domain (Clark et al., 2007; Laguri et al., 2008). The PWWP domain belongs to the Royal superfamily, which also includes the chromodomain and Tudor domain (Qin & Min, 2014). This domain possesses a conserved aromatic cage for histone methyl-lysine recognition, which binds H3K36me3-methylated nucleosomes in the context of DNA (Qin & Min, 2014).

Subsequently, several studies associated the PWWP interaction domain of hMutSα with the exonic enrichment of histone 3 lysine 36 trimethylation (H3K36me3) (Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009). *In vivo* pull-down assays demonstrated that a PWWP-mediated H3K36me3 interaction was necessary to localise hMutSα to chromatin (Li et al., 2013), and disrupting the H3K36 trimethyltransferase *SETD2* caused both increased microsatellite instability and increased mutation rate genome-wide (Li et al., 2013). Furthermore, a ChIP-seq analysis of hMSH6 found that 97% of hMSH6-enriched genes overlapped with H3K36me3 enrichment (Huang et al., 2018). This supports a model where chromatin states associated with actively transcribed genes recruit MMR factors and thereby reduce mutation rates within genes (Huang et al., 2018).

Interestingly, an analysis of mutations occurring in ENCODE cell lines revealed that somatic mutation rates within exons were reduced, in comparison to introns, based on the expected mutation rate given prior knowledge of how sequence composition effects mutation rate

(Frigola et al., 2017). The effect of sequence context was calculated based on the trinucleotide context of observed mutations occurring in a panel of previously sequenced tumours, which was used to calculate the expected mutation rate at each of the 192 possible trinucleotide contexts (Frigola et al., 2017). Exonic H3K36me3 levels negatively correlated with mutation rate for several cancer cell lines (Frigola et al., 2017). To validate this conclusion, glioblastomas were sequenced from children with inherited biallelic MMR deficiency caused by mutations in either *MLH1*, *MSH2*, *MSH6* or *PMS2*, and found that exonic mutation rates did not differ from the random expectation (Frigola et al., 2017). In summary, this study strongly suggests that MMR activity is non-randomly distributed across genes, and concentrated over coding sequences (Frigola et al., 2017).

In Arabidopsis, a genome-wide analysis of ~9,000 *de novo* mutations accumulating during five generations of self-fertilised *msh2-1* mutants revealed a disproportionate number of mutations falling within genes, compared to the wild type (Belfield et al., 2018). This suggests that MMR preferentially protects genic regions (Belfield et al., 2018). A recent study that systematically profiled Arabidopsis histone readers identified a Tudor domain in MSH6, a domain also known to interact with methylated histones (Huang et al., 2006; Zhao et al., 2018). Arabidopsis MSH6 weakly interacts with H3K36me3, but strongly binds to histone 3 lysine 4 di- or trimethylation (H3K4me2 or H3K4me3), a histone modification associated with gene transcription and enriched at the 5'-end of gene bodies (Zhao et al., 2018). As a plant-specific feature of MSH6, this suggests the existence of H3K4me2- or H3K4me3-mediated associations between MMR proteins and chromatin in proximity to genes, which may potentially alter the mutation landscape across the genome.

Furthermore, the nucleosome remodelling potential of MMR was demonstrated *in vitro*, using hMSH2-hMSH6 and reconstituted nucleosomes assembled on a synthetic *Xenopus laevis* DNA duplex (Javaid et al., 2009). The DNA duplex contained a 5′ 5S rDNA nucleosome positioning sequence, followed by a 28 bp linker, a 24 bp *lacO* sequence, a 47 bp sequence with or without a G/T mismatch, and a biotin-labelled 3′ DNA end. The latter feature prevents hMutSα's release from the 3′ duplex end when bound by streptavidin. An *Exo*III protection assay revealed nucleosome disassembly to be dependent on the presence of the DNA mismatch and ATP hydrolysis, which stimulated the release of hMutSα as a sliding clamp. hMutSα-induced nucleosome disassembly was boosted by the nucleosome destabilising histone modification H3K115 or H3K122 acetylation, whilst the binding of lacI to the *lacO* sequence separating the mismatch and the nucleosome binding domain repressed nucleosome disassembly (Javaid et al., 2009). These observations led the authors to propose that iterative loading of sliding clamps pushes nucleosomes away from mismatches (Javaid et

al., 2009), and implicates MMR activity in modifying the local chromatin environment (Javaid et al., 2009; Hauer & Gasser, 2017).

In addition, MSH2-dependent nucleosome eviction was also observed using a plasmid-based assay in *Xenopus* egg extracts (Terui et al., 2018). Upon addition of perfectly paired naked plasmid DNA, nucleosomes were deposited and induced plasmid supercoiling (Terui et al., 2018). However, the presence of a single G/T mismatch caused nucleosomes to be excluded from a ~1 kb region surrounding the mismatch (Terui et al., 2018). Notably, nucleosome exclusion was not dependent on Mlh1, suggesting a MutSα/β-specific role in remodelling (Terui et al., 2018). Pre-assembled nucleosomes were also excluded by this reaction, which required the Smarcad1 chromatin remodelling ATPase (Terui et al., 2018). Together, these *in vitro* and *in vivo* studies suggest that nucleosome eviction is a general feature of MMR (Javaid et al., 2009; Terui et al., 2018). Therefore, although initially preloaded onto chromatin via their interactions with nucleosomes, the initiation of MMR may cause subsequent nucleosome eviction.

1.3 The role of MMR in homologous recombination

In addition to DNA mismatches produced due to biosynthetic errors, when homologous recombination occurs between divergent sequences there is the potential to generate DNA mismatches (fig. 7). Here, I will review the literature relating sequence divergence and MMR to recombination. I will try to highlight the scale and type of polymorphism being assayed and the form of recombination outcome. A special focus will be given to both the role of *MSH*-homologs, as well as the relationship between recombination, polymorphism and MMR in plants. This survey will begin with bacterial systems, and progress towards 'higher' eukaryotes. Finally, the role of interhomolog polymorphism and MMR in regulating meiotic recombination genome-wide will be discussed.

1.3.1 A review of bacterial studies

In *Escherichia coli*, studies of phage integration and plasmid transformation indicated that sequence divergence repressed recombination (Méjean & Claverys, 1984; Shen & Huang, 1985; Watt et al., 1985). *In vivo* recombination between phage lambda and a plasmid vector revealed recombination to be linearly dependent on the length of homologous sequences, and reducing homology to 90% identity caused recombination frequency to reduce over 40-fold (Shen & Huang, 1985). Monitoring recombination frequency between insulin sequences, when contained in a plasmid and bacteriophage X vector, revealed that ~20 bp of homology was sufficient for recombination, and that increasing homology up to 300 bp caused a linear increase in frequency (Watt et al., 1985). In this system, a single variant between 53 base pair repeats was shown to decrease recombination frequency 4-fold (Watt et al., 1985).

Interestingly, recombination reactions between the related phage ΦX174 and G4 DNA sequences, diverging by 33% in their coding regions, efficiently catalysed joint molecule formation in the presence of the RecA recombinase (DasGupta & Radding, 1982). The observation that joint molecules could form in the presence of mismatches implicated the anti-recombination mechanism in acting downstream of strand invasion, and suggested that MMR activity may play a role in the inhibition (Karimova et al., 1985). This idea was extended by Miroslav Radman (1989) who proposed that MMR activity was anti-recombinogenic whenever mismatches were formed by strand exchange between non-identical parental sequences.

Support for this hypothesis was provided by mutagenesis screens for hyper-recombination phenotypes based on Hfr × F- bacterial crosses, which identified five *mutS* alleles and one *mutL* allele (Feinstein & Low, 1985). In a plasmid-phage recombination assay, the *mutS* mutant increased recombination rate when the homology of a recombinational substrate decreased from 100% to 89% (Shen & Huang, 1989). For example, whilst recombination rate decreased 240-fold in the wild type, recombination only decreased 9-fold in the *mutS* mutant (Shen & Huang, 1989). The largest increase was observed with *uvrD* (DNA helicase II) mutants, which increased recombination during Hfr × F- crosses by 20-fold (Feinstein & Low, 1985), implying that MMR directed anti-recombination required helicase activity. Notably, the degree of mutator phenotype did not correlate with the degree of hyper-recombination phenotype for these mutations (Feinstein & Low, 1985), suggesting that the mechanism of mismatch-dependent anti-recombination may not correspond to replication-associated MMR.

Conjugal recombination between a series of bacterial isolate crosses revealed that divergence repressed recombination, and the extent of recombination inhibition was found to be a function of sequence divergence (Vulić et al., 1997). Strikingly, it was also demonstrated that the barrier to recombination during conjugal or transductional crosses between *E. coli* and *Salmonella typhimurium* was relaxed in MMR mutants, where recombination increased by 1,000-fold (Rayssiguier et al., 1989). Hence, MMR acts as a form of ‘species’ barrier in bacteria.

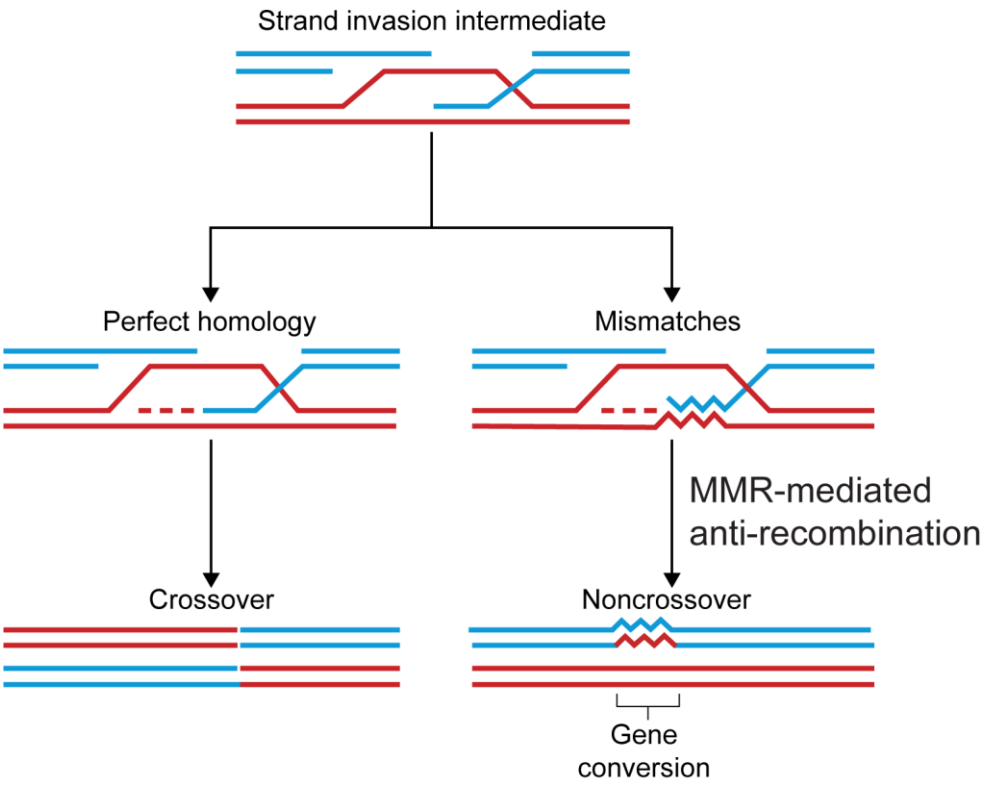


Figure 7. Mismatches are generated during recombination between divergent sequences. Adapted from Ziolkowski & Henderson (2017).

Schematic representation of meiotic recombination with or without interhomolog polymorphism, with homologous DNA strands shown in red and blue. In the absence of polymorphism, no mismatches are formed during strand invasion and crossovers form via Class I (interfering) or Class II (noninterfering) pathways. In the presence of interhomolog polymorphism, mismatches can be generated during strand invasion or branch migration, potentially leading to MMR-mediated anti-recombination. This results in dissolution of the D-loop and potentially leads to a gene conversion event. Gene conversions occur due to sequence amplification from the homologous chromosome and subsequent reannealing, followed by mismatch generation and MMR.

An *in vitro* characterisation of MutS and MutL proteins in bacteria showed them to abolish RecA-mediated strand invasion between fd and M13 bacteriophage DNA (Worth et al., 1994). Crucially, this abolition depended on mismatches between 3% diverged sequences, as strand exchange between homologous M13-M13 or fd-fd substrates continued unimpeded (Worth et al., 1994). By analysing strand-exchange intermediates accumulating in the fd-M13 reactions, the authors inferred that MutS and MutL inhibited branch migration, implying that recombination junctions may become stalled at sites of DNA mismatch (Worth et al., 1994). It was also demonstrated that overexpressing MutS and MutL *in vivo* strengthened the negative

relationship between mismatches and recombination (Vulić et al., 1997). Notably, deleting the C-terminal 53 amino acids of MutS (*mutSΔ800*) resulted in loss of anti-recombination activity, despite retaining mutation avoidance activity (Calmann et al., 2005). This observation indicates that MMR and anti-recombination in bacteria may act through distinct mechanisms (Calmann et al., 2005). However, although recombination between 17% diverged sequences was increased by up to 283-fold in the *mutSΔ800* background, this was only half the increase seen for a larger deletion (*mutSΔ680*), leading the authors to argue that this separation-of-function effect was in fact the consequence of reduced activity, rather than a loss of anti-recombination activity *per se* (Calmann et al., 2005). However, these observations have implications for understanding the effects of MMR proteins on anti-recombination genome-wide during meiotic recombination, given the potential for a large number of mispaired recombination intermediates to form and the implication that a large number of MMR proteins would be required to mediate anti-recombination activity.

1.3.2 A review of budding yeast studies

Budding yeast has been a major eukaryotic model species for genetic and biochemical studies of MMR and homologous recombination (Heyer et al., 2010; Chakraborty & Alani, 2016). An early study investigated whether meiotic crossovers are also accompanied by gene conversion events, by studying recombination within a 9 kb chromosome interval containing the DSB hotspot *URA3* (Borts & Haber, 1987). This interval was flanked by *MATa* and *MATα* mating type loci, which showed a recombination frequency of 22% during meiosis (Borts & Haber, 1987). Strikingly, introducing seven restriction site polymorphisms within the interval caused a 50% reduction in crossover frequency, indicating that interhomolog polymorphism impedes crossover repair (Borts & Haber, 1987). However, monitoring gene conversion across the interval indicated no reduction in total recombination events (Borts & Haber, 1987). The fact that total recombination levels were maintained suggested that introducing polymorphisms into recombination substrates in meiosis altered the pathway(s) used to resolve recombination intermediates, rather than DSB frequency. Indeed, the authors suggested that MMR events in the context of a recombination intermediate may create additional DSBs, or possibly an extensive length of ssDNA, that would be active substrates for further recombination (Borts & Haber, 1987).

Consistent with previous studies in bacteria (Rayssiguier et al., 1989), a two-dimensional gel analysis of branched recombination intermediates within an IC21 diploid strain, which carries one budding yeast chromosome III and one *Saccharomyces carlsbergensis–S. cerevisiae* mosaic chromosome III, revealed that no recombination occurred between the hybrid regions (Collins & Newlon, 1994). Furthermore, it was observed that sterility and chromosome

nondisjunction phenotypes in hybrids of budding yeast and *Saccharomyces paradoxus* were alleviated by mutating either *PMS1* or *MSH2* (Hunter et al., 1996). For example, crossover frequency at four intervals was reduced by ~10-75-fold in the hybrids, but increased by ~6-16.5-fold and ~2.5-10-fold in the *msh2* and *pms1* mutants relative to the wildtype, respectively (Hunter et al., 1996). The limited increases in MMR mutants suggested either additional anti-recombination mechanisms, sequence divergence *per se*, or other hybrid incompatibilities create a barrier to meiotic recombination between divergent genomes (Hunter et al., 1996).

The role of MMR in repressing somatic HR has been well characterised in budding yeast (Saparbaev et al., 1996; Datta et al., 1997; Selva et al., 1997; Welz-Voegele & Jinks-Robertson, 2008; Mitchel et al., 2010). For example, during somatic recombination, a single mismatch was found to repress recombination between inverted repeats, which depended on MMR components (Datta et al., 1997). In agreement with the limited restoration of recombination between highly diverged hybrid chromosomes in MMR-deficient backgrounds (Hunter et al., 1996), a 350 bp substrate was refractory to recombination beyond a sequence divergence of 10% with or without MMR activity (Datta et al., 1997). This again suggested that mismatches destabilise recombination intermediates independently of MMR proteins.

In order to disentangle the influence of MMR on crossover versus non-crossover outcomes, a plasmid integration system was used to monitor somatic recombination (Welz-Voegele & Jinks-Robertson, 2008; Mitchel et al., 2010). In this system, somatic recombination was induced by gap-based repair after DSB induction (Welz-Voegele & Jinks-Robertson, 2008). Non-crossover repair produced *URA3*-containing circular plasmids and hence an unstable Ura^+ colony phenotype, whilst somatic crossovers stably integrated the *URA3*-containing plasmid (Welz-Voegele & Jinks-Robertson, 2008). Interestingly, HR rate was decreased by ~2-fold in both the *msh2* and *msh3* mutants, whilst the *msh6*, *mlh1* or *pms1* did not alter total recombination rates (Welz-Voegele & Jinks-Robertson, 2008), suggesting that MMR proteins potentially play a role in mediating HR.

Overall recombination rates were reduced 2.4-fold in the gap-based repair assay when recombination substrates were diverged by 2% (Welz-Voegele & Jinks-Robertson, 2008). Notably, crossovers were reduced 7.3-fold, whilst non-crossovers were only reduced 1.4-fold (Welz-Voegele & Jinks-Robertson, 2008), perhaps indicating that anti-recombination acts post-strand invasion. This reduction was partially rescued by the *msh2* mutation, and to a lesser extent by the *msh3* mutation (Welz-Voegele & Jinks-Robertson, 2008). Interestingly, combining the same gap-based repair assay with sequencing of the recombination products revealed that crossovers associate with more extensive heteroduplex DNA (hDNA) tracts, compared to non-crossovers (Mitchel et al., 2010). This finding potentially explains the greater

repressive effect of MMR proteins on the mitotic crossover pathway, as there is more hDNA generated and hence more mismatches available for detection, perhaps generated by more extensive branch migration. In summary, studies of somatic recombination in budding yeast have revealed both a positive role for MMR proteins in mediating HR DNA transactions between homologous sequences, whilst also acting to repress HR between divergent sequences.

Using a chromatin immunoprecipitation approach, Evans et al. (2000) provided the first evidence for MMR localising to recombination structures *in vivo*. After crosslinking, HA-tagged Msh2 was broadly associated with genomic DNA, which depended on the presence of either Msh3 or Msh6 (Evans et al., 2000). After inducing HO-nuclease generated DSBs on a plasmid substrate, Msh2 associated strongly with the nonhomologous ends, suggesting that Msh2 may play a role in the homology search (Evans et al., 2000). In their system, mutating Rad50 reduced Msh2 association, suggesting that exonucleolytic processing is required for Msh2 binding (Evans et al., 2000). Together, this analysis provided strong biochemical evidence for the direct involvement of MMR in the somatic recombination process *in vivo*.

Filter-binding experiments established that budding yeast Msh2 binds to synthetic Holliday junctions with a 25-fold greater affinity than heteroduplex DNA containing a single mismatch (Alani et al., 1997; Marsischky et al., 1999). Competition binding experiments also demonstrated that interactions with Holliday junctions were more stable than those with heteroduplex DNA (Alani et al., 1997). To address the specificity of the protein-DNA interaction, electron microscopy was performed on Holliday junctions with extended 565-570 bp arms. Strikingly, 82% of dHJs with bound Msh2-Msh6 heterodimer showed localisation to the junction centre (Marsischky et al., 1999). The signal size indicated that proteins bound as either tetramers or larger oligomeric complexes (Alani et al., 1997). These observations suggested MSH2 to have an intrinsic ability to recognise Holliday junctions, as well as mismatches (Alani et al., 1997). Furthermore, Msh2-Msh6 binding promoted *in vitro* HJ cleavage by the T4 endonuclease VII and T7 endonuclease I (Marsischky et al., 1999). Together, these studies suggest that MMR proteins associate with recombination intermediates, and can putatively promote HJ cleavage. These results also connect with earlier studies that found mismatch recognition proteins to impede branch migration through mismatched sequences (Worth et al. 1994; Selva et al., 1997), and suggest that an association of MMR proteins with recombination intermediates may be a general feature of recombination, even in the absence of mismatches.

In summary, genetic and biochemical analyses of MMR proteins during HR in budding yeast reveal a complex relationship between polymorphism and recombination. Indeed, multiple lines of evidence suggest that budding yeast MMR proteins promote homologous

recombination, whilst repressing recombination between divergent sequences. However, Chakraborty & Alani (2016) developed a molecular model for MMR-directed anti-recombination, incorporating recent studies demonstrating the association of MMR proteins with chromatin (fig. 8) (see section 1.2.5), which raises the question of whether this model is applicable to meiotic recombination.

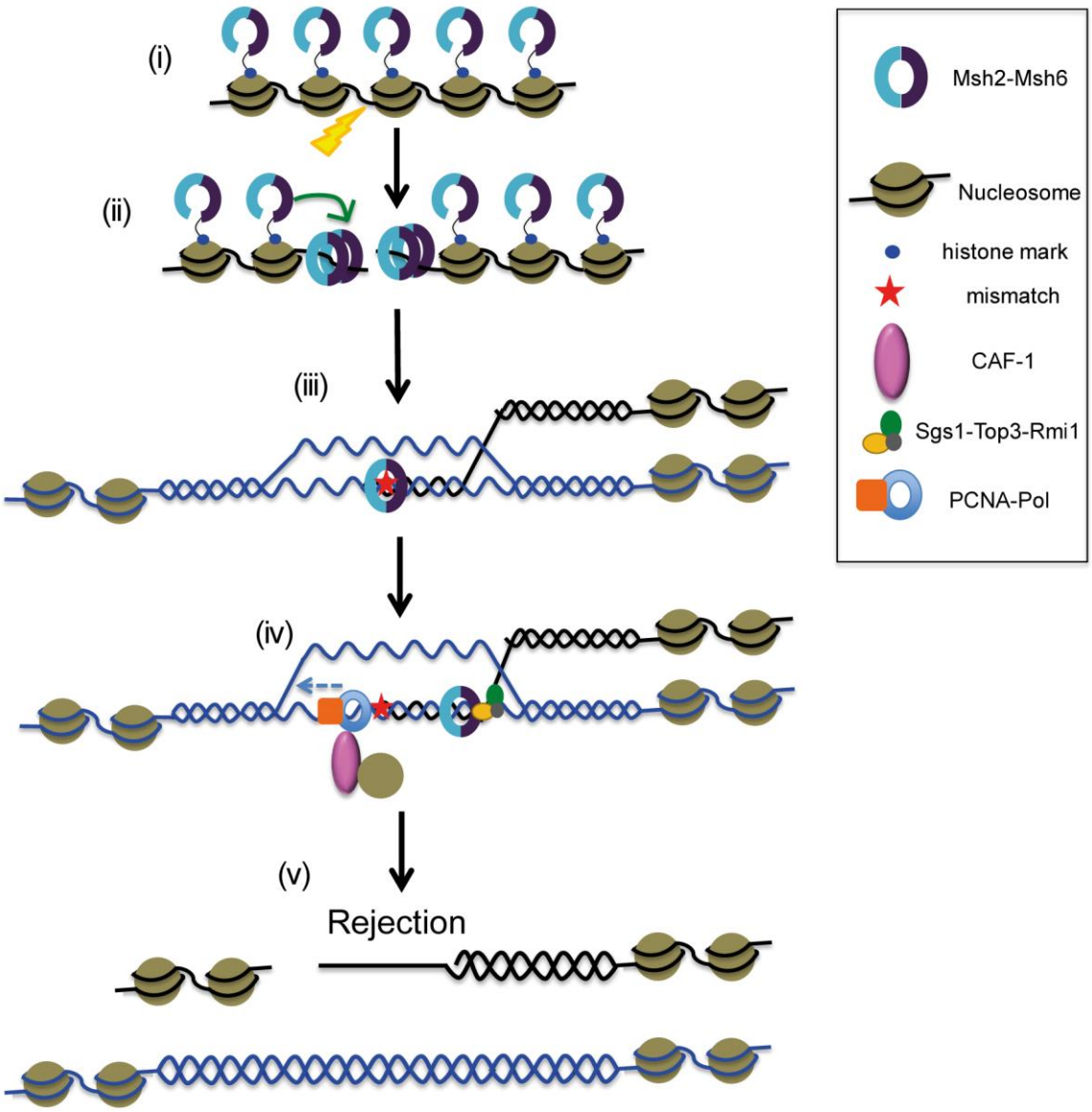


Figure 8. A molecular model for MMR-directed anti-recombination. Adapted from Chakraborty & Alani (2016).

MMR proteins (e.g. Msh2-Msh6) are associated with nucleosomes via their interactions with histone modifications, prior to DSB formation (i). Msh2-Msh6 can then rapidly associate with the DSB ends (ii), in order to recruit anti-recombination factors if mismatches (red star) are generated during strand invasion (iii). Therefore, the extent of prior Msh2-Msh6 association with chromatin determines the efficiency of MMR-directed anti-recombination. As DNA synthesis progresses, nucleosome remodellers such as CAF1 interact with PCNA, leading to synthesis-coupled nucleosome deposition (iv). In this model, nucleosome deposition inhibits anti-recombination, and stabilises the recombination intermediate. However, if anti-recombination activity is fast enough, the strand invasion is rejected (v).

1.3.3 A review of mammalian studies

Multiple studies investigating gene targeting in mammalian ES cells observed sequence divergence to impede recombination (de Wind et al., 1995; Elliott et al., 1998; Elliott & Jasin, 2001; LaRocque & Jasin, 2010). For example, a gene targeting assay in ES cells showed that a sequence divergence of 0.6% caused a 50-fold reduction in integration efficiency, whilst this divergence had no effect in the *msh2* background (de Wind et al., 1995). However, it should be noted that the absolute frequency of homologous gene targeting was also reduced by ~35% in the *msh2* background (de Wind et al., 1995), an observation later confirmed (LaRocque & Jasin, 2010). Furthermore, gene targeting efficiency was still decreased by divergence in a *msh2* mutant, showing that MMR is not the only barrier to recombination between diverged sequences (Elliott et al., 1998). Interestingly, unlike budding yeast, the BLM helicase (the mammalian *Sgs1* ortholog) was not required for supressing HR between diverged sequences (LaRocque & Jasin, 2010). Together, these results are consistent with MMR antagonising recombination between divergent sequences, and perhaps playing a role in mediating some aspect(s) of the recombination reaction.

In contrast, studies of endogenous meiotic crossover hotspots in mouse give a more complex picture (Cole et al., 2010). The *A3* meiotic crossover hotspot contains 32 polymorphisms within a 2 kb region in A/J and DBA/2J hybrids (1.6% divergence), and has a crossover frequency several hundred-fold greater than the genome average (Cole et al., 2010). Both crossover and non-crossover events can be identified using sperm-typing, an assay where crossovers or non-crossover molecules can be selectively amplified from sperm DNA (Cole et al., 2010). In contrast to the strong inhibitory effect of sequence divergence on somatic recombination, meiotic recombination was high within this interval (Cole et al., 2010). However, whilst gene conversions were present across the interval, crossovers were supressed in a region associated with structural polymorphism (Cole et al., 2010). Interestingly, estimates of

crossover frequency based on pedigree analysis in less divergent hybrids were similar to those of the more diverged hybrids (Kelmenson et al., 2005; Cole et al., 2010). For example, *Mus musculus* × *Mus musculus spretus* hybrids, which have a sequence divergence of 2.4%, had a higher crossover frequency than the A/J and DBA/2J hybrids (Kelmenson et al., 2005; Cole et al., 2010).

Kolas et al. (2005) performed an immunolocalisation study of MSH2 localisation in meiosis, using inbred mouse lines. In spermatocytes, MSH2 localised to repetitive centromeric satellite sequences in a MSH3-dependent manner, perhaps indicating a role for MutSβ in mediating meiotic recombination between repetitive DNA (Kolas et al., 2005). In addition, MSH2 foci were observed at low levels across the meiotic chromosome arms, consistent with a more general role for MMR in meiotic recombination (Kolas et al., 2005), even in the absence of interhomolog polymorphism.

1.3.4 A review of plant studies

Somatic and meiotic recombination have been widely studied in plants (Puchta & Hohn, 2012). A co-transfection plasmid assay in *Nicotiana plumbaginifolia* protoplasts monitored recombination between truncated *GUS* repeats on separate plasmids, via the restoration of β-glucuronidase (GUS) activity (Puchta & Hohn, 1991). This study found a positive correlation between homology length and recombination rate up to lengths of 1.2 kb, above which the rate of recombination plateaued (Puchta & Hohn, 1991). In this assay, reducing homology below ~450 bp strongly repressed recombination (Puchta & Hohn, 1991). This study demonstrated that plants, like other eukaryotes, require a minimum degree of sequence homology for efficient HR.

In Arabidopsis, intra-chromosomal somatic recombination has been frequently assayed using a 'split *GUS*' reporter, where the transgene contained two overlapping regions of the *GUS* coding sequence (*'GU'* and *'US'*, where *'U'* indicates a repeated central region) separated by a linker region (Puchta & Hohn, 2012). Recombination between the *'U'* repeats restores β-glucuronidase activity (Li et al., 2004b). Increasing the '*U'* repeat size from 153 to 589 bp caused a 183-fold increase in recombination rate, whereas a sequence divergence of 1.9% between the repeats caused a 10-fold reduction in recombination (Li et al., 2004b). Site-directed mutagenesis of the *'U'* repeat introduced increasing sequence divergences of 0.162%, 0.485%, 0.971%, or 1.618% (Opperman et al., 2004). Interestingly, the magnitude of recombination rate reduction was greatest with the introduction of a single polymorphism, and beyond 0.971% divergence there was no additional increase in recombination suppression (Opperman et al., 2004). These data contrast with budding yeast, where almost no mitotic HR can occur beyond a sequence divergence of 1% (Datta et al., 1997), indicating that barriers to

somatic HR caused by sequence divergence may be less stringent in plants. Together, these studies demonstrate a conserved role for sequence divergence in repressing somatic recombination in plants.

A number of studies in Arabidopsis have established the anti-recombination activity of MMR during somatic recombination (Li et al., 2006; Emmanuel et al., 2006). For example, relative to the wild type, there was a 2- to 7-fold increase in somatic recombination between diverged *'U'* repeats in the *msh2-1* background (Li et al., 2006). In contrast to budding yeast (Chen & Jinks-Robertson, 1999), this increase did not correlate with divergence level (Li et al., 2006). One possible explanation for this discrepancy is that budding yeast assays are typically performed in rapidly dividing mitotic cells, whereas somatic recombination in plants is typically studied in quiescent somatic tissues (Li et al., 2006; Emmanuel et al., 2006). This is consistent with the observation that Arabidopsis *MSH2* transcripts are more strongly detected by Northern blotting and hybridisation analysis in rapidly dividing cell cultures, and not in leaf tissue (Adé et al., 1999). Furthermore, mismatches still repress somatic HR within split *GUS* reporter constructs in the *msh2-1* background, implicating additional factors in mismatch-dependent anti-recombination (Li et al., 2006).

MMR has also been shown to play an anti-recombination function in more basal plant species (Trouiller et al., 2006). For example, in the moss *Physcomitrella patens*, a somatic HR gene targeting assay demonstrated that a sequence divergence of 1, 2, or 3% could potently supressed recombination in this system (Trouiller et al., 2006). However, in the *P. patens msh2* background sequence divergence no longer repressed gene targeting frequency (Trouiller et al., 2006). Notably, the *P. patens msh2* mutation also reduced the overall gene targeting efficiency of a homologous template by 2-fold (Trouiller et al., 2006). This was also observed for HR within the non-diverged *GUS* repeats in Arabidopsis, which decreased by 22% in *msh2-1* mutants, relative to the wild type. A similar phenomenon has also been observed with budding yeast and mouse somatic recombination assays (de Wind et al., 1995; Selva et al., 1997; Chen & Jinks-Robertson, 1999; LaRocque & Jasin, 2010), and is particularly intriguing as it supports a role for MMR in positively regulating somatic HR.

MMR has also been shown to impact meiotic recombination in plants (reviewed in Ziolkowski & Henderson, 2017). In Arabidopsis, meiotic recombination at a single 2 Mb sub-telomeric interval on chromosome 5, flanked by RFP or GFP transgenes with seed-specific expression, was monitored in a Col/Ler F_1 hybrid (Emmanuel et al., 2006). This experimental system utilises the natural variation between Arabidopsis accessions in order to create interhomolog polymorphism within the defined interval. Crossover frequency was measured by assessing the segregation of linked hemizygous fluorescent reporter transgenes in seed, as a crossover

within the interval splits the transgenes and produces GFP or RFP single-colour gametes (Melamed-Bessudo et al., 2005; Ziolkowski et al., 2015). In this system, loss of *msh2* function caused a 40% increase in meiotic crossover frequency, relative to the wildtype hybrid (Emmanuel et al., 2006). The authors concluded that MSH2 acts as a meiotic anti-crossover factor in hybrid plants (Emmanuel et al., 2006). However, given that this study analysed only a single reporter interval, it is unclear the extent to which these phenotypes may extend genome-wide (Emmanuel et al., 2006; Ziolkowski et al., 2015).

Several investigations of recombination at the maize *bronze* (*bz*) crossover hotspot found interhomolog polymorphism to influence meiotic recombination (Dooner, 1986; Dooner & Martínez-Férez, 1997; Dooner, 2002; Dooner & He, 2008). The *bz* locus provides an elegant system for studying recombination, as rare intra-locus recombination events are visible due to changes in seed pigmentation (Dooner, 1986). The presence of an insertion mutation in the middle of the *bz* locus resulted in a contraction of the *bz* genetic map, compared to alleles containing only single nucleotide polymorphisms (SNPs), indicating that insertions suppress meiotic crossovers to a greater extent than SNPs (Dooner, 1986). Measuring meiotic recombination between two *bz* progenitor alleles, differing at 21 positions across the 1,521 bp coding region, also showed that recombination was not evenly distributed, but biased towards regions of lower polymorphism (Dooner & Martínez-Férez, 1997). This trend was confirmed in a later study, where meiotic crossover junctions were mapped within the *bz* locus (Dooner & He, 2008), using two different hybrid combinations (McC × W22 & McC × B73). For both hybrids, plotting the frequency of crossovers and polymorphisms across the 6.7 kb interval clearly demonstrated a negative correlation between crossover frequency and polymorphism density (Dooner & He, 2008).

The observations in maize are also consistent with the negative relationship between crossover frequency and polymorphism at Arabidopsis crossover hotspots, as measured using pollen-typing (Drouaud et al., 2013; Serra et al., 2018b). This method is analogous to mammalian sperm-typing, and involves the selective amplification of crossover or parental molecules from pollen DNA derived from Col/Ler hybrid plants, using allele-specific PCR (Kauppi et al., 2009; Cole et al., 2010; Drouaud & Mézard, 2011; Choi et al., 2016). Crossovers can be distinguished due to the exchange of markers flanking the region of interest (Drouaud & Mézard, 2011; Choi et al., 2016). Deep sequencing of crossover molecules enables the fine-scale crossover landscape to be analysed (Serra et al., 2018b). At two Arabidopsis resistance gene crossover hotspots, *RAC1* and *RPP13*, crossovers were shown to anti-correlate with SNP density within the locus analysed (Serra et al., 2018b). Therefore, interhomolog polymorphism shapes crossover distributions at the fine-scale.

1.3.5 Genome-wide studies of meiotic recombination and MMR

As discussed, extensive studies have investigated interactions between MMR, polymorphism, and both mitotic and meiotic recombination. However, it is also noteworthy that many of these studies are based on analyses of single genetic intervals. Hence, it is unclear to what extent these findings can be extrapolated to patterns of meiotic recombination throughout the genome. For instance, meiotic recombination frequency and distribution are the product of a series of hierarchical factors, with many complex feedback loops, such as DSB feedback (Thacker et al., 2014; Garcia et al., 2015) and crossover interference (Copenhaver et al., 2002; Ziolkowski et al., 2015). Furthermore, while the anti-recombination activity of MMR has been clearly established for somatic recombination between diverged repeats, is remains unclear how this context, where a cell is facing a small number of HR events at any given time, relates to that of meiosis, where a cell faces potentially hundreds of recombination events. Indeed, in contrast to somatic recombination, a highly stringent MMR-directed anti-recombination system in meiosis could have deleterious effects for outcrossing species with a high level of genetic variation and as a consequence a high degree of mismatched joint molecules following interhomolog strand invasion. For example, if MMR was highly active in suppressing meiotic recombination, the high incidence of mismatched intermediates experienced during interhomolog recombination could cause crossover rates to fall below a critical threshold required for balanced homolog segregation and fertility.

A further point of consideration is that many studies of MMR-directed anti-recombination have relied upon ectopic reporter constructs, which may not be representative of the endogenous genome, or on recombination hotspots within the genome itself. To improve the power and/or efficiency of the experiment, the choice of construct or locus is typically selected to have a higher than average recombination frequency, whether endogenous (e.g. Cole et al., 2010) or synthetic (e.g. Borts & Haber, 1987), making these loci potentially non-representative. A further additional concern is that many assays involve ectopic DNA substrates, such as pre-formed mismatched plasmids introduced to mouse ES cells (Modrich, 2006). It is unclear how such studies relate to anti-recombination acting during meiosis, where mismatches are thought to be generated during the dynamic processes of strand invasion, second-end capture, and/or branch migration, all generated in the context of a complex chromatin environment. Finally, it is also important to consider that MMR during meiosis must operate in the context of a meiotic chromosome structure, as well as in the context of specific pro-crossover factors (e.g. the ZMM pathway). This raises questions about how relevant mechanisms of somatic HR are to meiotic recombination.

1.3.6 Sequence diversity, meiotic recombination and mutation rates

Polymorphism density is widely appreciated to vary across the genome (Cutter & Payseur, 2013). Interestingly, studies in multiple eukaryotes have revealed a positive relationship between historical levels of meiotic crossover and levels of natural genetic diversity, e.g. SNPs (reviewed in Cutter & Payseur, 2013). For example, an analysis of linkage disequilibrium in 19 Arabidopsis accessions observed a positive correlation between historical recombination rate and SNP density, when partitioning the genome into 200 kb windows (r^2 = 0.32; Kim et al., 2007). Also, linkage disequilibrium within a Eurasian Arabidopsis population peaked at the 5′ of genes, coinciding with a peak in local SNP density (Choi et al., 2013). Disentangling this relationship is challenging, because mutation rate may be elevated at sites of frequent meiotic recombination (Magni & Von Borstel, 1962; Strathern et al., 1995; Rattray et al., 2015). For example, the budding yeast *CAN1HIS3* forward mutation reporter was placed at a crossover coldspot or hotspot, and recombination was assessed in random spores selected with or without a mutation in *CAN1* (Rattray et al., 2015). In this system, recombination rate correlated positively with mutation rate, and was dependent on Spo11-induced meiotic DSBs (Rattray et al., 2015). These studies highlight the challenges to studying the role of interhomolog polymorphism in shaping the recombination landscape genome-wide, and also point to intriguing positive correlations between diversity and meiotic recombination rate in Arabidopsis.

1.3.7 Megabase-scale studies of interhomolog polymorphism and MMR in meiotic recombination

Regardless of these caveats, studies in a range of systems have experimentally addressed the relationship between interhomolog polymorphism and meiotic recombination genome-wide, and the role of MMR in this process. For example, a high resolution map of meiotic crossovers in a budding yeast S96 (an S288C derivative) and YJM789 hybrid, with a median distance of 78 bp between markers, showed an average crossover frequency of 90.5 crossovers and 46.2 non-crossovers per meiosis (Mancera et al., 2008). Surprisingly, an average genome-wide sequence divergence of ~0.6% had little effect on crossover frequency, and the total number of crossovers was close to the 86 crossovers predicted for a homozygous budding yeast strain, based on a genetic map determined through tetrad analysis (Cherry et al., 1997). In addition, the crossover map in the S96/YJM789 hybrid identified known crossover hotspots, such as *HIS4* and *ARG4*, and broadly correlated with a Spo11-oligo DSB map from an inbred SK1 strain (Mancera et al., 2008). Together, the consistencies with DSB maps, total crossover number and crossover hotspots measured in less diverse or inbred strains suggests

that interhomolog polymorphism may play a minor role in shaping meiotic recombination at the genome-scale (Mancera et al., 2008).

More recently, a genome-wide tetrad analysis of meiotic recombination was performed using a budding yeast S288C/SK1 hybrid, with an average divergence of ~0.57% (Martini et al., 2011). The ~65,000 SNPs in this cross were evenly distributed over the genome, with 97.5% of SNP intervals less than 1,000 bp, at an average density of 1 per 194 bp (Martini et al., 2011). The seven wild type meioses analysed showed an average of 73 crossovers, whilst the three *msh2* mutant meioses showed a significant increase to 92 crossovers per meiosis (Martini et al., 2011). Therefore, the activity of MMR in wild type causes a ~20% reduction in crossovers genome-wide. Interestingly, spore viability in the hybrid was reduced to 70% in the wild type and 63% in the *msh2* background, relative to the inbred strains (Martini et al., 2011). The authors hypothesised that the reduction in spore viability is caused by the accumulation of recessive lethal mutations in the *msh2* mutant generation, and is not due to meiotic defects (Martini et al., 2011). Particularly noteworthy is a recent study in budding yeast, that assessed crossover frequency and distribution in a *msh2* mutant (Cooper et al., 2018). The *msh2* mutation was again found to increase crossover frequency in an S288C×SK1 hybrid, and alleviated the suppression of Class I crossovers at regions of high sequence divergence. This study will be discussed in more depth in section 6.3. Together, these studies show that MMR plays a direct anti-crossover role in determining meiotic crossover frequencies genome-wide.

Experiments mapping Arabidopsis crossovers genome-wide confirmed the population genetic observation that recombination is weakly correlated with polymorphism density (Cao et al., 2011; Long et al., 2013; Fernandes et al., 2018a; Serra et al., 2018a). For example, within the chromosome arms crossovers show a positive correlation with polymorphism density in a wild type Col/Ler hybrid (r = 0.564) (Serra et al., 2018a). Consistent with this broad-scale association, juxtaposing regions of heterozygosity and homozygosity during Arabidopsis meiotic recombination caused the remodelling of crossovers, with an increase in the heterozygous region at the expense of the homozygous region (fig. 9A) (Ziolkowski et al., 2015). This phenomenon, termed the '*cis* effect', was dependent on the interfering ZMM Class I crossover pathway, as remodelling was absent in a *zip4* mutant background (Ziolkowski et al., 2015). The authors hypothesised that crossover remodelling may occur due to delayed recombination progression in heterozygous regions, leading to the recruitment of crossover-promoting 'late' DSBs (Ziolkowski et al., 2015). This model is consistent with observations of DSB feedback during meiosis, such as the negative DSB interference mediated via ATM/ATR kinase signalling (Carballo et al., 2008; Zhang et al., 2011; Kurzbauer et al., 2012; Garcia et al., 2015; Lange et al., 2016), and the observation that DSB levels increase in budding yeast *zmm* mutants that have a reduced crossover frequency (Thacker et al., 2014). However, the

authors also speculated that the presence of mismatches themselves may cause increased crossovers, perhaps due to the interaction of MMR machinery causing recombination intermediates within heterozygous regions to become preferentially directed along crossover designation pathways (Ziolkowski et al., 2015), consistent with observations in maize (Dooner, 2002). Thus, crossover frequency in Arabidopsis appears positively associated with interhomolog polymorphism at the megabase-scale, suggesting a more complex relationship between diversity, meiotic recombination and MMR.

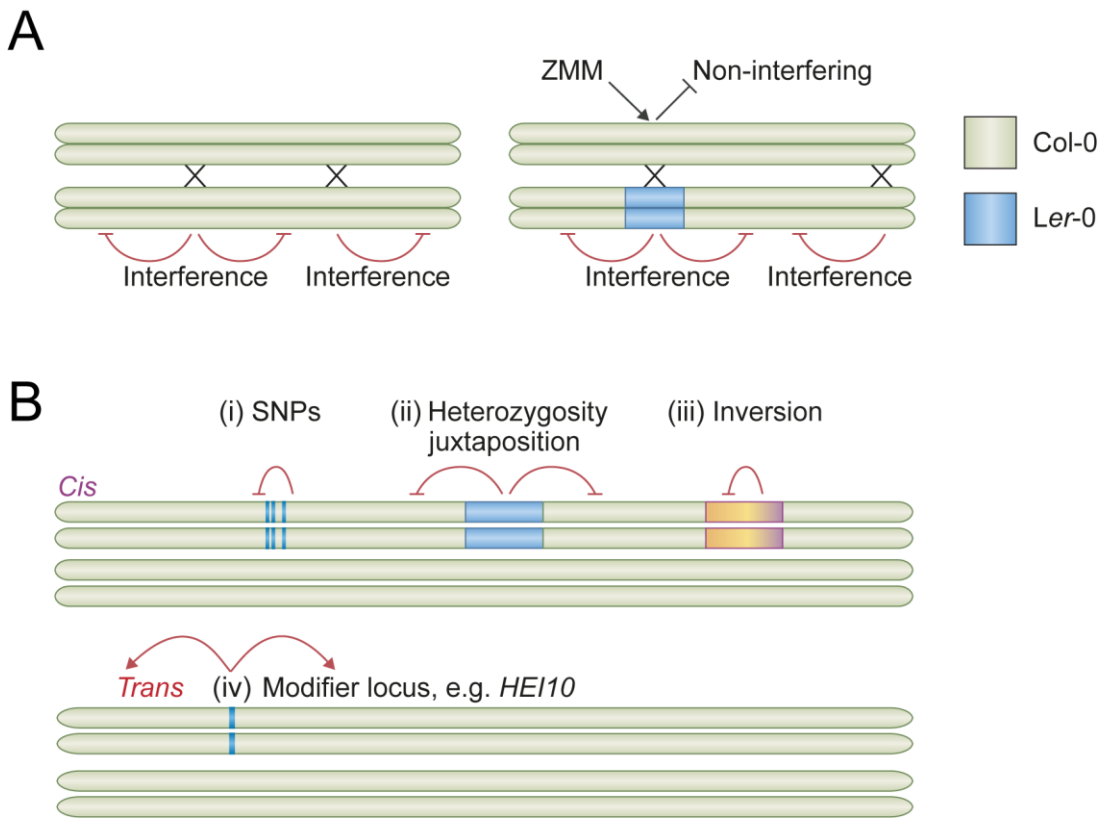


Figure 9. Interhomolog polymorphism shapes the meiotic recombination landscape. Adapted from Ziolkowski & Henderson (2017).

(A) Schematic representation of the *cis* effect. In the absence of polymorphism, crossover designation represses crossovers from formation in adjacent regions, due to crossover interference (left). Crossover remodelling occurs at megabase-scale when regions of homozygous and heterozygous sequence are juxtaposed, as heterozygous regions attract Class I crossovers and repress crossover formation in adjacent homozygous regions. In contrast, the ZMM pathway is inhibited by interhomolog polymorphism. (B) Interhomolog polymorphism influences meiotic recombination through a combination of *cis* and *trans* effects: (i) SNPs and indels repress local crossover formation, at the hotspot-scale, (ii) the *cis* effect (A), (iii) inhibiting crossovers due to large-scale genome rearrangements, such as inversions or duplications, and (iv) polymorphism leading to changes in a diffusible recombination modifier, such as HEI10, which can influence meiotic recombination both locally and globally.

However, an analysis of meiotic recombination in 17 Arabidopsis F_2 populations, derived from 18 different parental accessions, found no correlation between total genome polymorphism level and crossover frequency (Salomé et al., 2012). This observation is consistent with analyses of crossover frequency at several fluorescent reporter intervals (Barth et al., 2001; Ziolkowski et al., 2015), where recombination in hybrids was both increased and decreased relative to a Col-0 inbred, depending on the hybrid (Ziolkowski et al., 2015). Measurements of meiotic recombination in two maize populations also showed crossover rates to be highly variable between hybrids, but the authors observed a weak overall correlation with levels of polymorphism (Bauer et al., 2013; Rodgers-Melnick et al., 2015). Hence, it appears that within-genome variations in polymorphism density plays a greater role in shaping meiotic crossover distributions than total polymorphism levels.

However, meiotic recombination is modified by a combination of *cis* and *trans* acting modifiers of recombination (fig. 9B) (Lawrence e al., 2017). *Cis*-modifiers are genetic or epigenetic variants that have a local effect on the recombination landscape, whilst *trans* modifiers are genetic variants that effect the production of a diffusible molecule that can affect recombination across the genome (Lawrence et al., 2017; Ziolkowski et al., 2017). For example, a natural variant of the ZMM gene *HEI10* was found to be a *trans* modifier of crossover frequency in Arabidopsis, and could explain 56.9% of the variation in crossover frequency within a Col/Ler F_2 recombinant population (Ziolkowski et al., 2017). These observations raise the important issue of distinguishing between effects of interhomolog polymorphism on meiotic recombination in *cis*, and the effects those polymorphisms may confer on the meiotic proteome (Lawrence et al., 2017). The extent of interhomolog polymorphism used in most studies of Arabidopsis hybrids may therefore be at a level where *trans* modifiers play a greater role, which explains why total polymorphism level is a poor predictor of meiotic recombination frequency in Arabidopsis (Salomé et al., 2012; Ziolkowski et al., 2015).

It is also notable that the different meiotic recombination pathways in Arabidopsis are differentially sensitive to polymorphism. For example, the Class II crossover pathway in Arabidopsis seems highly sensitive to polymorphism (Girard et al., 2015; Ziolkowski et al., 2015; Fernandes et al., 2018a; Serra et al., 2018a). In an inbred background, the *fancm* mutation increases meiotic crossovers by 3-fold, as measured cytologically or using fluorescent reporter intervals (Crismani et al., 2012). However, the crossover increase in *fancm* was recently found to be repressed by genome-wide heterozygosity between the recombining chromosomes, as marker segregation analysis in an F_2 population derived from a *fancm* mutant Col/Ler hybrid was not significantly different from wild type (Girard et al., 2015). This is consistent with the reduced efficiency of Class II crossover formation in heterozygous regions observed when studying the *cis* effect (Ziolkowski et al., 2015). In contrast, a study of crossover

frequency in rice hybrids (Nipponbare × Dongjin), revealed the rice *fancm* mutation to increase crossover frequency by 2.3-fold (Mieulet et al., 2018). However, SNP density between these hybrids was only 1 per 11 kb, compared with the Col/Ler density of 1 per ~200 base pairs (Zapata et al., 2016). This may suggest a threshold sensitivity to interhomolog polymorphism in the *fancm* background (Mieulet et al., 2018).

Unlike *fancm*, mutations in the Arabidopsis AAA-ATPase *FIGL1*, a regulator of Class II crossovers, was less affected by genome-wide heterozygosity (Girard et al., 2015; Fernandes et al., 2018a). This indicates that, at least in the *fgl1* background, Class II crossovers can still occur between polymorphic chromosomes (Girard et al., 2015). Hence, the hybrid-specific crossover suppression in *fancm* must be produced through an effect specific to this genetic background (Girard et al., 2015). In support of FANCM and FIGL1 regulating Class II crossovers via distinct mechanisms, crossovers were synergistically increased in the *fancm figl1* double mutant (Girard et al., 2015).

Crossover frequency is also elevated in the Arabidopsis *recq4a recq4b* double mutant, in both Col-0 inbreds and Col/Ler hybrids (Séguéla-Arnaud et al. 2015; Fernandes et al., 2018a; Serra et al., 2018a). Likewise, overexpressing *HEI10* also caused crossover increases in both inbred and hybrid contexts (Ziolkowski et al., 2017; Serra et al., 2018a). However, although *recq4a recq4b* and *HEI10* increased crossovers in the chromosome arms by 4.1-fold and 2.3-fold, respectively, both genetic backgrounds only increased crossovers over the pericentromere by 1.1-fold (Ziolkowski et al., 2017; Serra et al., 2018a). Therefore, the pericentromeric and centromeric regions remain refractory to meiotic recombination in mutant backgrounds causing crossover increases in either the Class I or Class II pathways (Fernandes et al., 2018a; Serra et al., 2018a). This effect causes crossover increases in these backgrounds to be anti-correlated with SNP density, and it remains to be established whether this is due to a causal relationship, or correlations with other genomic parameters such as heterochromatin (Fernandes et al., 2018a; Serra et al., 2018a).

1.4 Aims & objectives

There remain several puzzles and outstanding questions regarding the role of interhomolog polymorphism and MMR in shaping the genome-wide meiotic recombination landscape in plants. Hence, my PhD project aims to directly address the impact of sequence divergence on meiotic crossover levels and distribution, and the role of MSH2 in mediating these effects. More specifically, this project aims to:

a) Characterise meiotic crossover frequency in the *msh2-1* mutant at several defined genetic intervals, in multiple hybrid backgrounds, thereby extending a previous analysis (Emmanuel et al., 2006).

b) Map crossover positions in wild type and *msh2* mutant populations via the low coverage sequencing of recombinant F_2 Arabidopsis individuals. This low coverage sequencing strategy reduces costs, and was adapted from a previously published protocol (Rowan et al., 2015). This experiment aims to assess the role of MSH2 in determining total crossover number and distribution, in relation to interhomolog polymorphism density. These effect of losing MMR activity on the relationship with polymorphism will be investigated at a range of physical scales: (i) at the fine-scale (1-50 kb scale) and (ii) the potential broad-scale redistribution of crossovers across the chromosomes (Mb scale).

c) Cytological and fertility phenotypes will also be investigated, to determine if losing MMR activity causes defects in meiotic chromosome behaviour.

In addressing these questions, the Arabidopsis model system provides several notable benefits. As a small flowering plant, Arabidopsis can be propagated from seed-to-seed in approximately six weeks, enabling multi-generation genetic experiments within the timescale of graduate research activity. In addition, its small size, ease of growth, high level of seed production from a single plant and relatively low outcrossing rate (~0.3%) make this model system excellent for addressing genetic questions in plant biology (Koornneef & Meinke, 2004). Arabidopsis is a diploid organism and the haploid genome is comprised of 5 chromosomes (AGI, 2000). The Arabidopsis genome is relatively small (~135 Mb), therefore being amenable to genome-wide sequencing experiments, and has many established data analysis pipelines (e.g. Lambing et al., 2019). Arabidopsis is a well characterised genetic system, with readily available genetic resources (Weigel, 2012). These include extensive mutant collections (e.g. Alonso et al., 2003), a well assembled and annotated genome sequence (AGI, 2000), large numbers of inbred natural accessions of Arabidopsis with high quality sequence variant maps and epigenomic datasets, such as the Col-0 and L*er*-0 accessions (Alonso-Blanco et al., 2016; Kawakatsu et al., 2016; Lang et al., 2016; Zapata et al., 2016), and the widespread use of Arabidopsis hybrids in studying meiotic recombination (Osman et al., 2011; Lambing & Heckmann, 2018). Furthermore, hybrids of most Arabidopsis accessions are fertile and are easily crossed to generate F_1 hybrids in order to study the effects of interhomolog polymorphism.

In parallel, my PhD project also aims to establish a meiotic chromatin immunoprecipitation-sequencing (ChIP-seq) protocol for the study of meiotic proteins in Arabidopsis. More specifically, this project aims to:

a) Characterise the genome-wide distribution of MSH4, a member of both the ZMM class of pro-crossover proteins and the MSH family (Higgins et al., 2004). Although previous cytological analyses have visualised MSH4 as punctate foci along meiotic chromosomes (Higgins et al., 2004), the distribution of MSH4 along the chromosome axis remains unknown. The development of a ChIP-seq protocol for MSH4 will provide a tool for addressing this question.

b) Perform a fine-scale bioinformatic analysis of MSH4 enrichment in relation to genomic features, such as gene bodies, transposons and transcription.

Together, these investigations will further our understanding of MMR in regulating meiotic recombination in a complex eukaryotic genome, and will provide a high-resolution picture of the distribution of MSH4 to complement cytological studies in Arabidopsis (Higgins et al., 2004). Given the deep evolutionarily conservation of these pathways, this project will provide useful insights into meiotic recombination in Arabidopsis that can be compared with our knowledge in other systems. Such a comparative approach will facilitate the improved understanding of meiotic recombination in eukaryotes more generally.

Chapter Two – Materials & methods

2.1 Plant & bacterial methods

2.1.1 Plant material

The *msh2* (*msh2-1*, SALK_002708) T-DNA insertion line was obtained from the Nottingham Arabidopsis Stock Centre (NASC), and was previously characterised by Hoffman et al. (2004). Genotyping of *msh2-1* was performed by PCR amplification using msh2-F (5'-AGCGCAATTTGGGCATGTCT-3') and msh2-R (5'-CCTCCCATGTTAGGCCCTGTT-3') oligonucleotides for the wild type allele, and msh2-F and LBb1.3 (5'-ATTTTGCCGATTTCGGAAC-3') oligonucleotides for the *msh2-1* allele. The *msh4* (*msh4-1*, SALK_136296) allele was also obtained from NASC, and was previously characterised by Higgins et al. (2004). Genotyping of *msh4-1* was performed by PCR amplification using msh4-F (5'-CGGCTTCACTGCATCTATCTC-3') and msh4-R (5'-TGGAATGGATCAATGAGTTCC-3') oligonucleotides for the wild type allele, and msh4-R and LBb1.3 (5'-ATTTTGCCGATTTCGGAAC-3') oligonucleotides for the *msh4-1* allele. The Arabidopsis accessions Col-0, L*er*-0 and Ct-1 were kindly provided by Dr Charles Underwood, Dr Donna Bond and Dr Piotr Ziolkowski, respectively. A Cvi-0 line, with a substituted L*er*-0 chromosome 5 (CCCCL) and Col-0 introgressions on chromosomes 2 and 4 was obtained from Dr Catherine Griffin.

Fluorescent-tagged lines (FTLs) were used to report meiotic crossover frequency (Francis et al., 2007; Berchowitz & Copenhaver, 2008; Yelina et al., 2013). FTL lines are in a Col-0 *qrt1-2*/*qrt1-2* background and express fluorescent proteins driven from the pollen-specific *LAT52* promoter. Four FTLs, *I1b*, *I2f*, *I3c* and *I5a*, each varying in transgene location and interval size, were used in this study (table 1). These lines were kindly provided by Prof. Gregory Copenhaver (Francis et al., 2007; Berchowitz & Copenhaver, 2008; Yelina et al., 2013). Two Col-0 Traffic Line (CTL) reporter intervals, *5.10* and *5.11*, were kindly provided by Prof. Scott Poethig, which enable measurements of crossover frequency based on seed-expressed eGFP and dsRed fluorescent proteins driven by the *NapA* promoter (table 1) (Wu et al., 2015). The seed based fluorescent reporter interval *420*, spanned by T-DNAs expressing GFP and dsRed2, was kindly provided by Prof. Avraham Levy (table 1).

Interval	Chr	Method	T-DNA 1	T-DNA 2	Mb	Location
I1b	1	Pollen	3,905,442-eYFP	5,755,618-dsRed2	1.85	Interstitial
I2f	2	Pollen	18,286716-dsRed2	18,957,093-eYFP	0.67	Sub-telomeric
I3c	3	Pollen	3,126,994-eYFP	4,319,513-dsRed2	1.19	Sub-telomeric
I5a	5	Pollen	18,164,269-dsRed2	23,080,567-eYFP	4.92	Sub-telomeric
420	3	Seed	256,516-GFP	5,361,637-dsRed2	5.11	Sub-telomeric
5.10	5	Seed	6,501,045-dsRed	13,229,304-eGFP	6.73	(peri)centromeric
5.11	5	Seed	6,501,045- dsRed	13,470,052-eGFP	6.97	(peri)centromeric

Table 1. Details of fluorescent reporter intervals used for measuring meiotic crossover frequency.

The interval name is indicated, followed by the chromosome, type of measurement system, the location of the two T-DNAs and their encoded fluorescent protein, the width of the interval in megabases (Mb) and the location relative to features of the chromosome landscapes.

2.1.2 Plant growth conditions & propagation

Plants were cultivated at 20 °C and 60 % humidity in controlled growth chambers, with long-day photoperiods (16 hour days and 8 hour nights). Light intensity was 150 µmol. All plants were germinated on standard commercial soil (F2) with vermiculite, at a density of 6-16 plants per 9 cm^2 pot (depending on application), following 4 days stratification at 4 °C.

Plants were manually crossed using fine dissecting forceps. Closed buds were emasculated and then cross pollinated immediately. Crosses were re-pollinated the following day, and seeds were collected approximately 10 days post-pollination. Unless otherwise stated, all other Arabidopsis lines were propagated by self-fertilisation.

2.1.3 Agrobacterium tumefaciens-mediated Arabidopsis transformation

Col-0 plants were transformed with *A. tumefaciens* using the 'floral dip' method, as previously described (Clough & Bent, 1998; Zhang et al., 2006). Briefly, an *A. tumefaciens* colony was used to inoculate a 5 ml Lysogeny Broth (LB) starter culture, containing appropriate antibiotics (e.g. rifampicin (25 µg / ml) and kanamycin (50 µg / ml)). After 48 hours, 1 ml of this culture was used to inoculate a 500 ml LB culture containing appropriate antibiotics. After reaching an OD 600 of 1.5 – 2.0 this solution was centrifuged at 4000 × *g* for 10 minutes at 4 °C to pellet the cells. Cells were re-suspended in 1 volume of 5 % sucrose solution (w / v). Prior to dipping, Silwet L-77 was added to a final concentration of 0.02 % (v / v). Plants were dipped for 30

seconds, drained, placed in plastic covers, and remained in the dark for 24 hours. Plants were then grown in standard conditions.

2.1.4 Selection of Arabidopsis transformants

Seeds were sterilised with chlorine gas for 18 hours, and sown onto 1 % agar supplemented with Murashige and Skoog basal medium (50 % v / v), timentin (20 µg / ml), and nystatin (100 µg / ml). Plants were selected with kanamycin (50 µg / ml) or (10 µg / ml) glufosinate ammonium. Seeds were stratified at 4 °C for 4 days, before being germinated in controlled conditions (section 2.1.2). Transformants were transferred to soil at 10 days.

2.1.5 Pollen-based measurements of crossover frequency

Pollen-expressed fluorescent reporter transgenes (FTLs) were used to measure crossover frequency (Berchowitz and Copenhaver, 2008). Reporters were under the control of the pollen-specific *LAT52* promoter (Francis et al., 2007), and expressed either dsRed2 (R) or eYFP (Y) fluorescent proteins. When in both a hemizygous and *cis*-configuration (*RY*/++), the reporter T-DNAs will segregate during meiosis to give an altered pattern of fluorescence in the pollen depending on the presence of crossovers (fig. 10). Measuring the relative classes of fluorescence – either parental (double or non-colour) or recombinant (single colour) – enables the calculation of crossover frequency in centimorgans per Mb (cM/Mb) within the interval spanned by the transgenes. As the stock FTL lines are homozygous for the *qrt-2* mutation, causing the four pollen grains from a single meiosis to remain associated as tetrads and therefore become intractable to flow cytometry analysis (Yelina et al., 2013), the plants were phenotyped to select against this genotype after crossing.

To extract pollen, inflorescences were collected in 50 ml falcon tubes from mature plants with a *RY*/++ genotype and either a *QRT1*/*qrt1-2* or *QRT1*/*QRT1* genotype. Pollen Sorting Buffer (PSB; 10 mM $CaCl_2$, 1 mM KCl, 2 mM MES, 5 % sucrose (w / v), 0.01 % Triton X-100 (v / v), pH 6.5) was added, and the pollen extracted by vigorous shaking. The solution was filtered through a 40 µm cell strainer (Stemcell Technologies) into a fresh falcon tube and centrifuged at 450 × *g* for 5 minutes at 4 °C. The supernatant was gently discarded, the pellet washed with PSB (minus Triton X-100) and again centrifuged at 450 × *g* for 5 minutes at 4 °C. The supernatant was discarded and the pellet re-suspended in 600 µl PSB. Flow cytometry was performed on a BD Accuri™ C6 Flow Cytometer (BD Biosciences), equipped with a 488 nm laser and 530/30 nm and 570/20 nm band-pass filters. To select pollen by size, events were separated based on forward and side scatter. Hydrated pollen were gated to exclude dead or damaged material. Finally, events could be identified by emission signal into red (R3), yellow (R6), double-colour (R4) or non-colour (R5) categories (fig. 10). cM were calculated as previously described (Ziolkowski et al., 2015):

cM = 100 × (R6 / (R6 + R4))

Events from recombinant and non-recombinant classes were pooled to test for statistically significant differences between genotypes using χ^2 tests on 2×2 contingency tables (Ziolkowski et al., 2015).

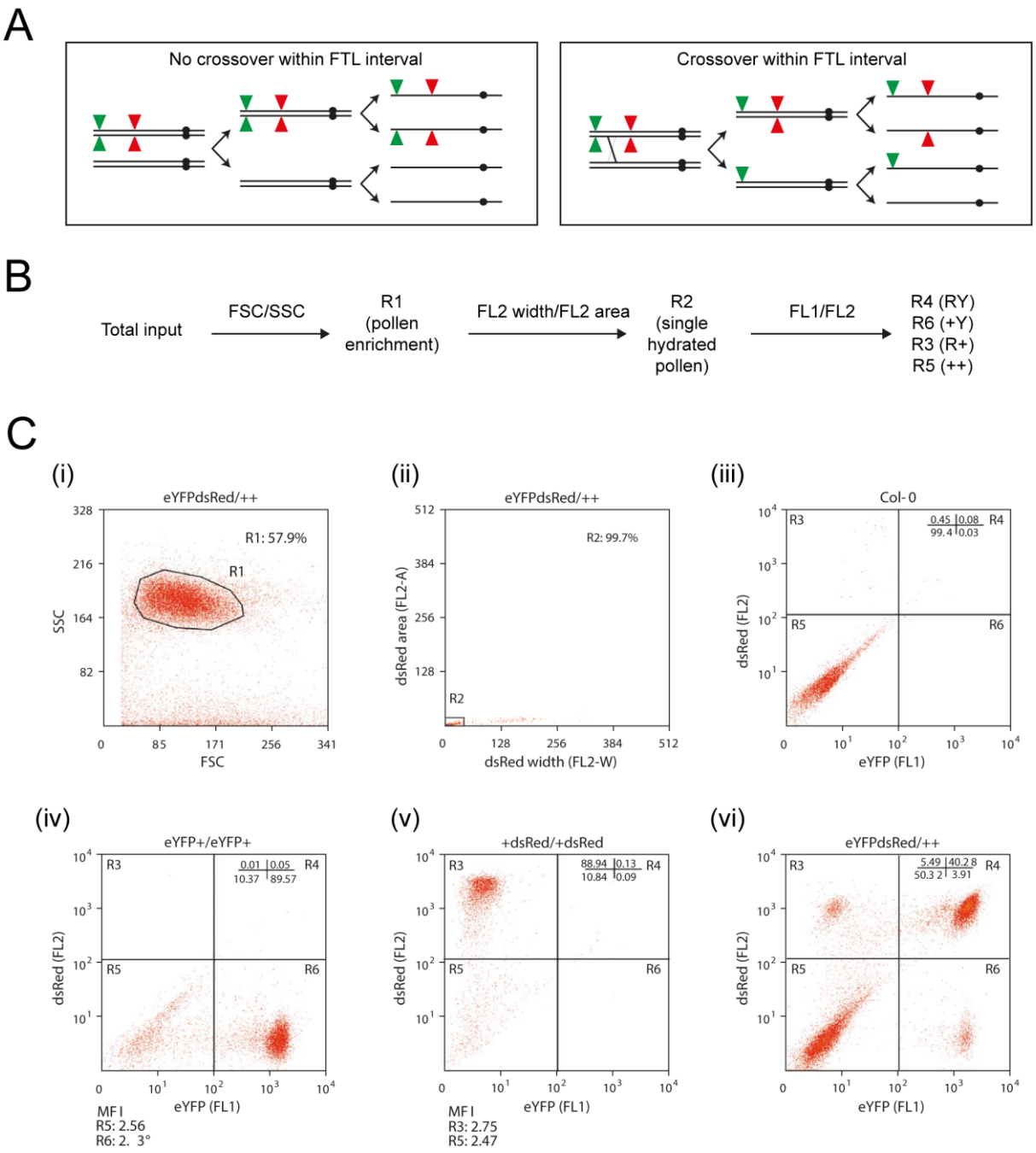


Figure 10. Flow cytometry analysis of crossover frequency using fluorescent FTL reporters. Adapted from Yelina et al. (2013).

(A) A schematic representation of reporter transgene segregation using the FTL system. If no crossover occurs between the linked hemizygous transgenes (red or green triangles), then half of the gametes will be non-colour and half will contain both red and green fluorescent markers (double-colour) (left panel). If a crossover (black cross) occurs within the interval, then one non-colour, one double-colour, one single-colour red and one single-colour green gamete will be produced (right panel). The four chromatids of a homologous chromosome pair are represented with black lines and centromeres are represented as black circles. The arrows represent the first round (left) and second round (right) of chromosome segregation in meiosis. (B) A schematic representation of the flow cytometry analysis pipeline. Pollen is analysed as bulk input, before using the forward scatter (FSC) to side scatter (SSC) ratio to gate the pollen objects (R1). Hydrated, single pollen are then gated (R2) by assessing the dsRed (FL2) width to FL2 area ratio. The eYFP (FL1) intensity of this subset is then plotted against FL2 intensity, revealing four classes of recombinants (R4, R6, R3 and R5). (C) Example FTL flow cytometry acquisition plots, showing (i) the selection of pollen objects into gate R1, (ii) the selection of single, hydrated pollen into gate R2, (iii) a control experiment with Col-0 pollen, showing that only non-fluorescent pollen is present (R5), (iv) a control experiment with *+Y*/*+Y* pollen, showing that only the eYFP fluorescent class is present (R6), as well as residual dead pollen (R5), (v) as for (iv) but with *R+*/*R+* pollen, and (vi) analysis of fluorescence classes from *RY*/++ segregating pollen, showing the four recombinant classes (R3, R4, R5, and R6).

2.1.6 Seed-based measurements of crossover frequency

Crossover frequency was measured by assessing the segregation of fluorescent reporter transgenes in seed (Melamed-Bessudo et al., 2005; Ziolkowski et al., 2015). Fluorescent reporters expressed either eGFP (G) or dsRed (R) under the seed-specific *NapA* promoter (Wu et al., 2015). Seeds were collected from plants hemizygous for transgenes in a *cis*-configuration (*RG*/++) and, after cleaning to remove debris, seed was imaged as a single layer with approximately 1,500-2,000 seeds per image. For each sample, three images were captured with a Leica M165 FC dissecting epifluorescence microscope (Leica Microsystems), with a) brightfield, b) mCherry filter and c) GFP filter. Images were processed through a custom CellProfiler pipeline, which identifies seed boundaries and assigns each a fluorescent intensity value (Carpenter et al. 2006; Ziolkowski et al. 2015). Seed were assigned to one of nine possible fluorescence classes (++/++, *RG*/++ (*R+*/*+G*), *RG*/*R+*, *RG*/*+G*, *R+*/*R+*, *R+*/++, *+G*/++, *+G*/*+G*, *RG*/*RG*), and thresholds between fluorescence and non-fluorescence were set manually using fluorescence histograms (Ziolkowski et al., 2015). Crossover frequency was calculated using the formula:

$cM = 100 \times (1 - (1 - 2(R+G) / T) / 2)$

Where R is red only seeds, G is green only seeds and T is the total number of seeds (Ziolkowski et al., 2015). Note that this formula is only appropriate to F_2 seed collected by self-fertilization of a plant with *RG*/++ genotype. Events from recombinant and non-recombinant classes were pooled to test for statistically significant differences between genotypes using χ^2 tests on 2×2 contingency tables (Ziolkowski et al. 2015).

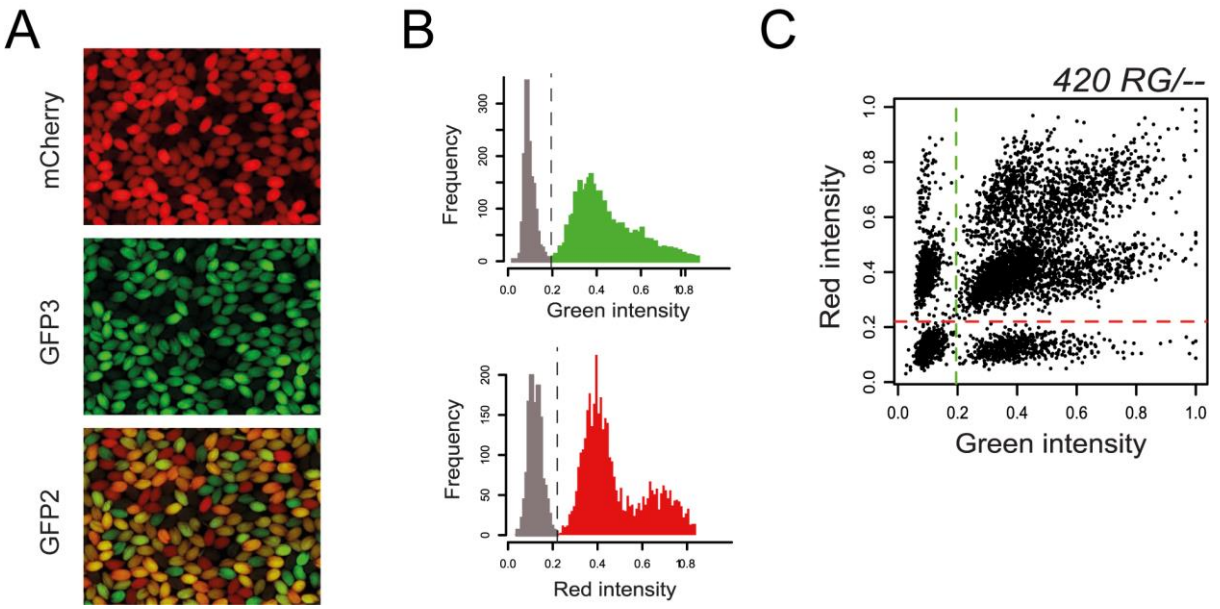


Figure 11. CellProfiler analysis of crossover frequency using fluorescent *420* or CTL seed-based reporters. Adapted from Ziolkowski et al. (2015).

(A) Example fluorescence micrographs of seed from segregating *420 RG*/++ individuals, imaged under a fluorescence microscope with mCherry, GFP3 or GFP2 filters. Images using the GFP2 filter enable all fluorescent classes to be distinguished. (B) Histograms of seed fluorescence intensity, generated using CellProfiler analysis. The coloured (green or red) and non-coloured (grey) fluorescence intensity threshold is marked with a vertical dashed line. (C) A scatter plot of individual seeds (black dots) plotted by red versus green fluorescence intensities, revealing the nine possible fluorescence classes. Red and green dashed lines mark the coloured/non-coloured threshold indicated in (B).

2.1.7 Escherichia coli strains

Chemically competent *E. coli* Subcloning Efficiency™ DH5α™ and One Shot® TOP10 strains were used for transformation and plasmid amplification. Both strains were supplied by Invitrogen™ (Life Technologies™). *E. coli* liquid cultures of LB medium were grown at 200 rpm at 37 °C overnight (16 hours). *E. coli* were also grown on LB agar plates containing the appropriate antibiotic. Plates were inoculated, inverted, and incubated at 37 °C for 20 hours.

2.1.8 Agrobacterium tumefaciens strains

The *A. tumefaciens* strain GV3101 was used for Arabidopsis transformation. Liquid cultures of *A. tumefaciens* were grown in LB medium containing rifampicin (25 µg / ml) and the appropriate antibiotic selection. Cultures were incubated at 200 rpm at 28 °C for 48 hours. *A. tumefaciens* were cultured on LB agar plates with rifampicin (25 µg / ml) and appropriate antibiotics. Plates were inverted and incubated at 28 °C for 2 – 3 days.

2.1.9 Escherichia coli transformation

E. coli DH5α transformations were carried out as follows. Plasmid DNA (pDNA) or assembly reactions were either diluted and added to 50 µl aliquots of competent cells, or added directly. The mixture was incubated for 30 minutes on ice, heat-shocked at 42 °C for 90 seconds, suspended in 500 µl SOC media, and grown at 37 °C for 60 minutes. Cells were plated on appropriate selection media. One Shot® TOP10 cells were transformed according to manufacturer's instructions.

2.1.10 Agrobacterium tumefaciens transformation

DNA constructs were electroporated into *A. tumefaciens*. pDNA was mixed with 50 µl of cells, and transferred to an electroporation cuvette (Bio-Rad). Electroporation was performed with a GenePulser Xcell™ (Bio-Rad) using the following parameters: 25 µF, 200 Ω, and 2400 V. 500 µl SOC medium was then added. The mixture was incubated at 200 rpm for 2 hours at 28 °C, and plated onto appropriate selection media.

2.2 Molecular biology methods

2.2.1 Bacterial plasmid extraction

All bacterial DNA extractions were performed with a QIAprep Spin Miniprep Kit (Qiagen), according to manufacturer's instructions.

2.2.2 Arabidopsis gDNA extraction for PCR genotyping

Genomic DNA (gDNA) for PCR genotyping was extracted as described by Edwards et al. (1991), but adapted for a 96-well format. Briefly, 200 µl of Extraction Buffer (200 mM Tris-Cl

pH 7.5, 250 mM NaCl, 25 mM EDTA) was added to leaf tissue, before disruption with a TissueLyser II (Qiagen). A further 200 µl of Extraction Buffer (with 1 % SDS) was added, and the solution was centrifuged at 300 rcf for seven minutes. The supernatant was transferred to a 0.8 ml storage plate and one volume of isopropanol was added. DNA was precipitated at room temperature for 10 minutes, before centrifugation at 3000 rcf for 35 minutes. The pellet was washed with 70 % ethanol and re-suspended in 150 µl of water.

2.2.3 Arabidopsis CTAB gDNA extraction

For cloning and genotyping-by-sequencing library preparation applications, high quality gDNA was extracted from three week old *Arabidopsis* plants. 3-4 leaves were collected into 2 ml Eppendorf tubes containing 4 glass beads (3 mm). Samples were snap frozen in liquid nitrogen and stored at -20 °C. Samples were again cooled to -80 °C before grinding in a TissueLyser II (Qiagen) for 3 rounds of 1 min (30 Hz). Samples were cooled in liquid nitrogen between each round. 700 µl of CTAB buffer (140 mM sorbitol, 220 mM Tris pH 8.0, 22 mM EDTA, 800 mM NaCl, 0.1 % (v / v) N-Lauryl sarcosine, 0.8 % (w/v) CTAB (cetyl trimethyl ammonium bromide)) was added to each Eppendorf, which were inverted until resuspension of the plant material was complete. Samples were incubated for 60 minutes at 65 °C with 700 rpm mixing in a Thermomixer (Eppendorf). Tubes were additionally inverted after 30 minutes. After cooling to room temperature, samples were briefly centrifuged to pellet insoluble debris. 650 µl of the supernatant was transferred to a fresh 2 ml Eppendorf containing an equal volume of chloroform (650 µl), and vortexed vigorously until mixed. Samples were then centrifuged at 13,000 rpm for 5 minutes at room temperature. 580 µl of the upper aqueous layer was removed to a new 1.5 ml Eppendorf containing an equal volume of isopropanol. Tubes were gently vortexed, left at room temperature for 4 minutes and then centrifuged at 13,000 rpm for 15 minutes at 4 °C. The supernatant was poured off and the pellet washed by centrifuging at 13,000 rpm with 500 µl of 70 % ethanol, at 4 °C. The supernatant was again poured off and the Eppendorf inverted to air dry for 20 minutes at room temperature. The pellet was re-suspended in 100 µl water containing RNase A (1 µl 100 mg / ml RNase A per 1 ml water) and incubated at 37 °C for 30 minutes. gDNA was precipitated by addition of 0.1 volumes of 3 M Sodium Acetate Solution pH 5.2 (11 µl) and 2.5 volumes of 100 % ethanol, and incubation at -20 °C for > 30 minutes. Eppendorfs were centrifuged at 13,000 rpm for 15 minutes at 4 °C. The supernatant was poured off and the pellet washed as described above. The supernatant was again poured off and the Eppendorf inverted to air dry for 20 minutes at room temperature. The final pellet was re-suspended in 25 µl of water.

2.2.4 Nucleic acid quantification

pDNA was quantified using a NanoDrop™ 1000 Spectrophotometer (Thermo Scientific). For accurate CTAB gDNA quantification, a BR DNA Qubit® Fluorometer (Life Technologies) was used. For NGS library quantification, a HS DNA Qubit® Fluorometer (Life Technologies) was used, in conjunction with an HS DNA Agilent 2100 Bioanalyzer system (Agilent Technologies).

2.2.5 Primer design & oligonucleotides used in this study

Unless otherwise stated, all primers were manually designed and assessed for Tm and secondary structure using OligoAnalyzer Tool (IDT) (https://idtdna.com/calc/analyzer). Cleaved Amplified Polymorphic Sequence (CAPS) and Simple Sequence Length Polymorphism (SSLP) markers were designed based on the SALK 1,001 Genomes Project polymorphism browser (Alonso-Blanco et al., 2016). SSLP markers were designed to show a size difference after PCR amplification between 35 bp and 300 bp in length. Lyophilized oligonucleotides were supplied by Sigma-Aldrich®. All primers used in this study are detailed in appendix 7.1 and appendix 7.2.

2.2.6 PCR amplification for T-DNA and marker genotyping applications

PCR amplifications were performed in 10 µl reactions using 1 U GoTaq® DNA Polymerase (Promega), 1× Green GoTaq® Reaction Buffer (Promega), 0.5 mM dNTPs (Bioline), 0.6 µM of forward and reverse primers, and varying quantities of genomic DNA. T-DNA genotyping reactions were performed with the following thermo cycles: [95 °C, 2 minutes] [95 °C, 30 seconds; 56 °C, 30 seconds; 72 °C, 1 minute 30 seconds] × 35 [72 °C, 3 minutes]. To assess the genotype at specific polymorphic sites across the genome, marker genotyping was performed with the following thermo cycles: [95 °C, 2 minutes] [95 °C, 10 seconds; 58 °C, 10 seconds; 72 °C, 40 seconds] × 35 [72 °C, 2 minutes].

2.2.7 High-fidelity PCR amplification for cloning applications

PCR products for cloning reactions were amplified in 20 µl reactions using 0.4 U Phusion High-Fidelity Polymerase (Thermo Scientific), in the presence of 1× Phusion HF Buffer (Thermo Scientific), 0.5 mM dNTPs (Bioline), 0.25 µM of forward and reverse primers, 3 % DMSO (Thermo Scientific) and varying quantities of gDNA or pDNA. Thermo cycles were according to manufacturer's instructions.

2.2.8 DNA gel electrophoresis & detection

Agarose gels were used to separate DNA molecules according to size. Agarose was dissolved in 1× TBE and boiled, with the agarose concentration depending on the size of DNA molecules to be separated. After cooling, ethidium bromide was added to a concentration of 0.4 µg / ml

and the solution was cast appropriately. For more sensitive applications, SYBR™ Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific) was added directly to the agarose gel to a 1× concentration.

2.2.9 DNA gel extraction & purification

After gel electrophoresis, PCR products and restriction digested DNA were visualised under long wavelength UV light. Bands were excised with a sterile scalpel and DNA was extracted using a QIAquick Gel Extraction Kit (Qiagen), according to manufacturer's instruction.

2.2.10 Epitope-tagged MSH4 construct assembly

An epitope linker and human influenza haemagglutinin (HA) tag were constructed from overlapping oligonucleotides and cloned into a pPZP-derived vector (kindly provided by Dr Kyuha Choi), upstream of a stop-codon and a *NOS*-terminator (Szakasits et al., 2007). This region was sequenced and cloned into the pGreenII-0229 vector using restriction enzymes *Pst*I-HF (NEB) and *Bam*HI-HF (NEB), and T4 DNA Ligase (NEB) according to manufacturer's instructions (Hellens et al., 2000). A Myc epitope tag was amplified from a pPZP-derived vector and similarly cloned into pGreenII-0229. All restriction enzymes and buffers were supplied by New England Biolabs. DNA restriction digests were performed according to manufacturer's instructions.

To create C-terminal tagged constructs, full-length genomic *MSH4* DNA was amplified from Col-0 gDNA including the native promoter (1.8 kb). pGreenII-0229 vectors containing HA or Myc tags were linearised with *Eco*RV-HF and fragments were assembled using Gibson Assembly® Master Mix (NEB) according to manufacturer's instructions (with a 50 °C incubation time of 2 hours). To synthesise N-terminal tagged constructs, the *MSH4* promoter and gene body were amplified separately and assembled sequentially into pGreenII-0229, with the promoter upstream of the tag, and the gene body downstream.

2.2.11 Sanger Sequencing

Sanger sequencing of pDNA was performed by Source Bioscience (http://www.lifesciences.sourcebioscience.com). Samples were provided according to company requirements.

2.2.12 ChIP-qPCR assays

qPCR was performed on 1:10 diluted ChIP DNA using a CFX96 Real-Time PCR machine (Bio-Rad) and GoTaq (Promega), in the presence of 0.5× SYBR™ Green (Thermo Fisher). The following thermo cycle was used: [95 °C, 3 minutes] [95 °C, 30 seconds; 56 °C, 30 seconds;

72 °C, 30 seconds + Plate Read] × 40 [Melt Curve 65 °C – 95 °C, increment 0.5 °C per 5 seconds + Plate Read].

2.2.13 NGS library preparation for genotyping-by-sequencing

Crossover locations were identified in a sample of 96 barcoded recombinant F_2 individuals by low-coverage (~4.5×) sequencing, using a protocol adapted from Rowen et al. (2015). CTAB gDNA samples were quantified with BR DNA Qubit® Fluorometer (Life Technologies) and normalised to 15 ng / µl in a 96-well plate. Sample quality was assayed by running 1.5 µl of all normalised samples on a 1 % agarose gel. 12 groups of 8 samples were selected, each group with matched band intensity, and were re-plated into a fresh 96-well plate. gDNA was sheared with 0.4 U of dsDNA Shearase Plus (Zymo Research) in a final volume of 15 µl. After incubation for 20 minutes at 37 °C, the reaction was terminated by a 5 minute incubation at 65 °C and the addition 1.6 µl of EDTA (0.5 M). Sheared DNA was cleaned up using 1.8 volumes of Agencourt AMPure XP beads (Beckman Coulter), according to manufacturer's instructions, and eluted in EB (10 mM Tris pH 8.0) in a final volume of 20 µl.

18 µl of eluted sample was end-repaired with 3 U of T4 DNA polymerase (New England Biolabs), 10 U of T4 polynucleotide kinase (Thermo Fisher Scientific), and 1.25 U of Klenow fragment (New England Biolabs) in the presence of 1 mM dNTPs in a reaction volume of 12 µl for 30 min at 20 °C. End-repaired DNA was cleaned up using 1.8 volumes of AMPure XP beads, according to manufacturer's instructions, and eluted in a final volume of 12 µl. To A-tail the DNA fragments, 10 µl of eluted sample was incubated for 30 minutes at 37 °C with 2.5 U of Klenow exo- (New England Biolabs) in the presence of 0.2 mM dATP, in a final volume of 15 µl. A-tailed DNA was cleaned up using 1.8 volumes of AMPure XP beads, according to manufacturer's instructions, and eluted in a final volume of 11 µl. The plate was briefly centrifuged to accumulate the AMPure XP beads in the base of the wells. To index each individual F_2 sample, custom Illumina adaptor barcodes were ligated using 1 µl of Quick Stick Ligase (Bioline) in a final volume of 20 µl. The reaction was incubated for 30 minutes at 20 °C and heat-inactivated for 10 minutes at 65 °C. Rows of 8 wells were combined into a 1.5 ml Eppendorf and 1.8 volumes of PEG solution (8 mM polyethylene glycol, 2.5 M NaCl) was added to rebind DNA to the AMPure XP beads. Pooled DNA was then cleaned up as described above, and eluted in a final volume of 30 µl.

To achieve an optimal library size (300 – 700 bp) fragments of > 700 bp were bound by the addition of 0.55 volumes of AMPure XP beads and incubated for 5 minutes at room temperature. Eppendorfs were placed on a magnetic rack for 5 minutes, until the supernatant had cleared. The supernatant, containing the desired DNA, was transferred to a new Eppendorf. To bind fragments > 300 bp, 0.23 volumes of AMPure XP beads were added and

incubated for 5 minutes at room temperature. DNA was cleaned up as described above and eluted in 20 µl. Libraries were amplified from 12 µl of eluted sample, with KAPA HiFi HotStart ReadyMix PCR Kit (Kapa Biosystems) and previously described primers (Rowan et al., 2015), according to the following thermo cycle: [95 °C, 3 minutes] [98 °C, 20 seconds; 65 °C, 30 seconds; 72 °C, 30 seconds] × 12 [72 °C, 5 minutes]. Amplified libraries were cleaned up using 1 volume of AMPure XP beads, according to manufacturer's instructions, and eluted in a final volume of 10 µl. Libraries were quantified according to section 2.2.4, diluted to 4 nmol, pooled in equal volume, and sequenced on an Illumina NextSeq 500 instrument using a NextSeq® 500/550 Mid Output Flow Cell (2 × 75).

2.2.14 ChIP-seq library preparation

To construct a ChIP-seq library for NGS, 10 ng of purified DNA was processed using a TruSeq Prep Kit v2 (Illumina), according to the following protocol. End-repair was performed on 10 ng of DNA with 5 U of T4 DNA polymerase (New England Biolabs), 50 U of T4 polynucleotide kinase (Thermo Fisher Scientific), and 5 U of Klenow fragment (New England Biolabs) in the presence of 0.4 mM dNTPs in a reaction volume of 100 µl for 30 min at 20 °C. The reaction was cleaned up using 1.8 volume of AMPure XP beads, according to manufacturer's instructions, and eluted in a final volume of 33 µl. A-tailing was performed by incubating 32 µl of end-repaired DNA at 37 °C for 30 minutes, with 15 U of Klenow exo- (New England Biolabs) in the presence of 0.2 mM dATP, in a final volume of 50 µl. The reaction was cleaned up using 1.8 volume of AMPure XP beads, according to manufacturer's instructions, and eluted in a final volume of 26 µl. Illumina adaptors were ligated to 25 µl of eluted sample by incubation at 20 °C for 30 minutes, using 5 µl of Quick Stick Ligase (Bioline) in a final volume of 50 µl. The reaction was cleaned up using 1 volume of AMPure XP beads, according to manufacturer's instructions, and eluted in a final volume of 20 µl. PCR amplification was performed with KAPA HiFi HotStart ReadyMix PCR Kit (Kapa Biosystems) using 19 µl of eluted DNA and 1 µl Universal Primer Mix (Illumina), in a total volume of 40 µl. PCR conditions were as follows: [95 °C, 3 minutes] [95 °C, 30 seconds; 60 °C, 30 seconds; 72 °C, 30 seconds] × 12 [72 °C, 5 minutes]. To achieve an optimal library size (250 – 370 bp) fragments of > 370 bp were bound by the addition of 0.6 volumes of AMPure XP beads and incubated for 5 minutes at room temperature. Eppendorfs were placed on a magnetic rack for 5 minutes, until the supernatant had cleared. The supernatant, containing the desired DNA, was transferred to a new Eppendorf. To bind fragments > 250 bp the DNA was cleaned up using 1 volume of AMPure XP beads, according to manufacturer's instructions, and eluted in a final volume of 15 µl. Sequencing was performed on a 4 nmol library dilution with an Illumina NextSeq 500 instrument using a NextSeq® 500/550 Mid Output Flow Cell (2 × 75).

2.2.15 Protein extraction

Protein was extracted from unopened *Arabidopsis* buds, for western blot analysis. Briefly, a 100 µl volume of floral buds was ground in liquid nitrogen, 100 µl of Protein Extraction Buffer (50 mM Tris-Cl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 10 % glycerol, 1 % Triton X-100, 1 mM PMSF, 1 mM DTT, 1× Roche cOmplete™ Protease inhibitor) was added, and the solution brought to 4 °C on ice whilst grinding. The solution was briefly vortexed, and then centrifuged at 13,300 rpm for 5 minutes at 4 °C. The supernatant was removed for quantification and western blot analysis.

2.2.16 Protein quantification

Protein extracts were quantified using the Qubit™ Protein Assay Kit (Thermo Fisher) according to manufacturer's instructions.

2.2.17 SDS-PAGE & western blotting

Protein extracts or eluted protein solutions were diluted in an equal volume of 2× SDS-PAGE Sample Buffer (0.16 M Tris-Cl pH 6.8, 4 % SDS, 20 % glycerol, 0.02 % Bromophenol Blue) and loaded on a 1.0 mm NuPAGE™ 3-8% Tris-Acetate Protein Gel (Invitrogen) in a NuPAGE Tris-Acetate SDS Running Buffer (Invitrogen). After resolution on the gel, protein transfer to an Immun-Blot® PVDF membrane (Bio-Rad) was achieved using a transfer solution (25 mM Tris pH 8.3, 192 mM glycine, 20 % v / v methanol) for 4 hours at 70 mV at 4 °C. Following transfer, the membrane was briefly rinsed with TBS-T (50 mM Tris pH 7.6, 150 mM NaCl, 0.1 % TWEEN® 20) and incubated overnight with Blocking Buffer (1× TBS-T, 5 % w / v non-fat Marvel Dried Skimmed Milk Powder) on a rocker at 4 °C. The Blocking Buffer was then discarded and the membrane briefly rinsed in TBS-T. A solution of primary antibody diluted in Blocking Buffer (α-HA: Roche, 3F10, dilution 1:3000; α-HA: abcam, ab9110, dilution 1:15000) was added and incubated for 2 hours at room temperature with constant agitation. Afterwards, the membrane was briefly rinsed in TBS-T and then washed at room temperature with TBS-T for 4 × 10 minutes with constant agitation. A solution of secondary antibody diluted in Blocking Buffer (α-rat IgG-HRP, Santa Cruz Biotechnology, sc-2006, dilution 1:8000; α-rabbit IgG-HRP, Santa Cruz Biotechnology, sc-2054, dilution 1:20000) was added and incubated for 1 hour at room temperature with constant agitation. Finally, the membrane was briefly rinsed in TBS-T and then washed at room temperature with TBS-T for 6 × 10 minutes with constant agitation. The signal was detected on Hyperfilm™ ECL (GE Healthcare Amersham™) using ECL™ Prime Western Blotting Detection Reagent (GE Healthcare Amersham™) and exposed using a Xograph Compact X4 X-ray Film Processor.

2.2.18 Immunoprecipitation of meiotic proteins

Since *MSH4* is specifically expressed in meiocytes (Higgins et al., 2004), unopened flower buds were collected for extraction and snap frozen in liquid nitrogen. These buds contain all stages of male and female meiosis. Approximately 1 gram of bud tissue was ground in liquid nitrogen in a pre-chilled mortar and pestle, and then transferred to a second mortar pre-chilled on ice. 5 ml of Lysis Buffer (25 mM HEPES-NaOH pH 7.9, 5 mM EDTA, 2 % SD, 1 mM PMSF, 2 mM DTT, 1× Roche cOmplete™ Protease inhibitor) was added and the sample ground until completely dissolved. The solution was transferred to a 50 ml falcon tube and incubated at 95 °C for 15 minutes in a water bath. The falcon was then immediately put on ice and the sample aliquoted into 1.5 ml Eppendorfs. After 13,000 rpm centrifugation at 4 °C, the supernatants were re-pooled and diluted with 3 volumes of IP Dilution Buffer (14 mM Tris pH 8.0, 1 % Triton X-100, 150 mM NaCl).

To pre-bind the antibody (α-HA: Roche, 3F10; α-HA: abcam, ab9110), 1 volume (50 µl) of Dynabeads™ Protein A (Thermo Fisher) was transferred to a 1.5 ml Eppendorf and placed on a magnetic stand for 30 seconds. The supernatant was removed and replaced with 10 µg of antibody diluted in 4 volumes (200 µl) of PBS-T (1× PBS, 0.02 % TWEEN® 20). The antibody-bead solution was incubated on a rotator for 20 minutes, at room temperature. The tubes were again placed on the magnetic stand for 30 seconds, the supernatant removed, and the antibody-bound beads re-suspended in PBS-T. Mock beads (i.e. without antibody) were processed in parallel.

To immunoprecipitate the MSH4 protein, 20 µl of antibody-beads (4 µg) was added per 3 ml of diluted sample, and the mixture incubated overnight at 4 °C. Beads were then removed from the solution using a magnetic stand, and washed briefly with 1 ml of IP Dilution Buffer. The solution was immediately removed and a further 1 ml wash performed. A further four washes with rotation at 4 °C for 10 minutes were then performed. The beads were gently re-suspended in 200 µl and transferred to a single Eppendorf tube. The tube was placed on a rack for 30 seconds and the supernatant removed. Proteins were eluted from the beads in 1× SDS-PAGE Sample Buffer (0.08 M Tris-Cl pH 6.8, 2 % SDS, 10 % glycerol, 0.01 % Bromophenol Blue) by incubating at 75 °C for 5 minutes, then a further 95 °C for 5 minutes. Samples were then kept on ice and stored at - 20 °C.

2.2.19 Chromatin Immunoprecipitation of meiotic proteins

Approximately 10 grams of unopened *Arabidopsis* buds were ground in liquid nitrogen using a pestle and mortar, in batches of 2 grams. Cross-linking was performed *in vitro* with 50 ml Nuclear Isolation Buffer (60 mM Hepes pH 8.0, 1 M sucrose, 5 mM KCl, 5 mM $MgCl_2$, 5 mM EDTA, 0.6 % Triton X-100, 0.4 mM PMSF, 1 mM pepstatin, 1× Roche cOmplete™ Protease

inhibitor, 1 % formaldehyde) per 2 grams of tissue, at room temperature for 25 minutes, with rotation. To quench the cross-linking reaction, glycine was added to a final concentration of 125 mM and the solution was incubated at room temperature for a further 25 minutes, with rotation. Nuclear isolation was achieved by filtering the solution twice through one layer of Miracloth (Merck Millipore), followed by centrifugation at 3,000 × *g* for 10 minutes at 4 °C. Further purification was achieved by gently re-suspending the nuclear pellet in EB2 (10 mM Tris-HCl pH 8.0, 1 mM EDTA , 0.25 M sucrose, 10 mM $MgCl_2$, 1 % Triton X-100, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 1 mM pepstatin, 1× Roche cOmplete™ Protease inhibitor), before centrifuging at 12,000 × g for 10 minutes at 4 °C. After discarding the supernatant, the pellet was re-suspended in 600 µl of Nuclei Lysis Buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1 % SDS, 0.1 mM PMSF, 1 mM pepstatin, 1× Roche cOmplete™ Protease inhibitor) and aliquoted equally into two Eppendorfs. DNA was sonicated using a Bioruptor (Diagenode) for 15 minutes (high intensity 320 W, [30 seconds ON; 30 seconds OFF]). Each 300 µl sample was then diluted with 200 µl of ChIP Dilution Buffer (20 mM Tris-HCl pH 8.0, 167 mM NaCl, 1.1 mM EDTA, 1.1 % Triton X-100, 1 mM pepstatin, 1× Roche cOmplete™ Protease inhibitor) and an additional cycle of sonication was performed for 25 minutes (high intensity 320 W, [30 seconds ON; 1 minute OFF]). Insoluble debris was then removed from solution by centrifugation at 5,000 × *g* for 10 minutes at 4 °C. The supernatant was incubated for 16 hours at 4 °C with 8 µg of antibody (α-HA, ab9110, abcam) pre-bound to 50 µl of Dynabeads™ Protein G (Thermo Fisher) per gram of input tissue. An aliquot (5 %) of the supernatant was snap frozen as a representative input sample.

After immunoprecipitation, the beads were washed twice with a Low Salt Wash (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1 % SDS, 1 % Triton X-100) on a rotator for 5 minutes at 4 °C, and a further two times with a High Salt Wash (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1 % SDS, 1 % Triton X-100). Bound chromatin was eluted in 200 µl of Elution Buffer (1 % SDS, 0.1 M NaHCO3) at 65 °C for ten minutes, with occasional vortexing. The supernatant was removed to a new tube and the elution step repeated with a further 200 µl of Elution Buffer. The eluates were combined to a final volume of 400 µl. Reverse crosslinking of DNA-protein complexes was performed overnight at 65 °C, in the presence of 240 mM NaCl. To digest RNA, 10 µl of 0.5 M EDTA, 20 µl of 1 M Tris pH 7.0, and 3 µl of RNase A (20 mg / ml) were added sequentially, and the solution incubated at 37 °C for 30 minutes. To digest protein, 2 µl of proteinase K (20 mg / ml) was added and the solution incubated at 45 °C for 2 hours. DNA was then precipitated and purified by standard phenol-chloroform extraction, and re-suspended in 50 µl of TE Buffer (10 mM Tris pH 8.0, 1 mM EDTA).

2.3 Cytological methods

2.3.1 Alexander's stain pollen viability assay

Mature flowers were selected from inflorescences on the primary floral axis, and anthers were agitated in 20 µl of Alexander's Stain Solution (0.01 % malachite green, 10 % ethanol, 0.05 % acid fuchsin, 0.005 % orange G, 4 % glacial acetic acid, 25 % glycerol) to release pollen grains. A cover slip was applied and sealed with a rubber solution (Weldtite). Slides were incubated overnight at 37 °C and screened for pollen viability using a standard brightfield microscope.

2.3.2 DAPI-stained chromosome spreads

Arabidopsis inflorescences were collected 3 weeks post-germination from the primary floral axis and fixed in Fixative Solution (3-to-1 ethanol-to-acetic acid) at 4 °C. The Fixative Solution was replaced after 3 hours, and again after 12 hours. Fixed inflorescences were dissected in fresh Fixative Solution using fine forceps under a Leica S6 E stereomicroscope (Leica Microsystems). Buds of length 0.2-0.7 mm were selected, corresponding to floral stages 8 to 10 (Smyth et al., 1990), which are known to contain a range of meiotic stages (Armstrong & Jones, 2003). The Fixative Solution was removed and the buds washed 3 times in 1 ml of Citrate Buffer (44.5 mM citric acid, 55.5 mM sodium citrate) for 2 minutes. To digest the cell walls, buds were incubated with an Enzyme Solution (3.3 mg / ml cellulose (Sigma) 3.3 mg / ml pectolysase (Sigma), diluted in Citrate Buffer) in a moist box for 1 hour 30 minutes at 37 °C. The reaction was terminated by removing the Enzyme Solution and adding 1 ml of Citrate Buffer. Individual buds were transferred to a drop of water on a glass slide, and gently disrupted with a brass rod. Following this, two 5 µl drops of 60 % acetic acid were added, the resulting solution mixed with a needle, and the slides incubated on a heat block at 48 °C for 1 minute. To fix the material, 150 µl of ice-cold Fixative Solution was applied to the slide, encircling the acid drop, and the slide rocked from side-to-side to spread the mixture over the central area of the slide. The slide was inverted and dried with a hairdryer. Finally, the chromatin was counterstained with 14 µl of DAPI Solution (10 µg / ml DAPI (Sigma), diluted in VECTASHIELD Antifade Mounting Medium). An inverted DMI6000 B microscope (Leica Microsystems) was used for screening and image capture of DAPI-stained slides.

3.3.3 DAPI spread heterochromatin quantification

Measurements of heterochromatin area at the pachytene stage on DAPI-stained slides were performed by Ms Pallas Kuo using ImageJ (NIH). Heterochromatic regions, excluding the nucleolus organiser regions, were selected based on relative pixel intensity value. The micrometre per pixel conversion parameter was defined by the DMI6000 B microscope (Leica

Microsystems) and used by ImageJ to measure the size of the defined heterochromatin regions.

3.4 Statistical & bioinformatics analyses

3.4.1 Genotyping-by-sequencing data analysis

For each accession, a high quality SNP dataset was generated by analysing all F_2 sequencing reads in order to call variants. Briefly, fastq files generated from the entire sequencing lane were aligned to the TAIR10 genome assembly using Bowtie2 using default parameters (version 2.2.4, Langmead & Salzberg, 2012). The SAM files produced were compressed to BAM format and then indexed and sorted using SAMtools (version 1.2). Variant sites were called using the BAM alignment files with SAMtools (version 1.2) and BCFtools (Li et al., 2009). Sites were then filtered to remove those with a quality score <100, >2.5× mean coverage, <10× reference coverage, or >100× variant coverage. Sites were repeat-masked as described previously (Choi et al., 2013).

To align each F_2 individual separately, fastq sequencing reads were demultiplexed using fastq-multx with a custom file containing 96 barcodes. Reads from each of the 96 individuals were aligned to the TAIR10 genome assembly using Bowtie2 using default parameters. The variant sites defined using the entire lane of data were compared to the aligned data from the de-multiplexed samples and calls obtained using SAMtools and BCFtools. For each library, the number of variant versus reference reads supporting each SNP site was obtained. These values were passed to TIGER (Trained Individual GenomE Reconstruction), a pipeline for reconstructing the mosaic genomes of recombinant individuals from low-coverage sequencing data (Rowan et al., 2015). Briefly, the genotype along each chromosome was estimated by counting read alignments at marker positions that supported either of the parental alleles (A or B) using a sliding window (Rowan et al., 2015). TIGER provides output files that give estimates of the coordinates of transitions between A/AB/B genotypes, which are taken as crossover positions. Crossover coordinates from all F_2 individuals were combined to generate a crossover map from the F_1 hybrid.

3.4.2 ChIP-seq data bioinformatic analysis

ChIP-seq data were analysed as previously described (Lambing et al., 2019). Paired-end (2×76 bp) ChIP-seq sequencing reads were deduplicated and aligned to the TAIR10 reference genome using Bowtie2 version 2.2.9 (Langmead & Salzberg, 2012), using the settings: --very sensitive --no-discordant --no-mixed -p 4 -k 10. A maximum of 10 valid alignments were reported for each read. Several filtering steps were then performed. Firstly, aligned reads with greater than 2 mismatches were discarded using the SAM optional field “XM:i”. Uniquely

aligning reads were then extracted by removing alignments with the SAM optional field “XS:i” and with Bowtie2-assigned MAPQ scores lower than 42. Multiple mapping reads were addressed by filtering against reads with MAPQ scores lower than 10, and from these reads only the alignment with the highest MAPQ score was retained. If MAPQ scores for multiple-mapping reads were equal, one alignment was randomly selected. Any alignments consisting of a single read within a read-pair were removed.

Filtered unique and multiple aligned reads were combined, and Rsamtools (version 1.12.1) was used to calculate coverage at each genomic coordinate. To control for sequencing depth variations, coverage was normalised by the total coverage of each library. Next, $\log_2$(ChIP/input) enrichment was calculated to control for background DNA and variation in mappability between genomic loci. A Z-score standardisation was then applied such that the genome-wide mean coverage equals zero, and a value of plus or minus one equates to one standard deviation from the mean. These values were used for subsequent analyses and data visualisations. HA-MSH4 ChIP-seq peaks were identified using ranger tool within the PeakRanger suite (Feng et al., 2011), with ChIP input DNA as a control background control. As HA-MSH4 is broadly associated with chromatin across the genome, stringent significance thresholds ($P \leq 0.001$ and FDR ≤ 0.01) were applied for the detection of peaks. This analysis produced a total of 40,556 HA-MSH4 peaks genome-wide

Several available datasets were utilised for comparison with the HA-MSH4 ChIP-seq data. These included a paired-end REC8 ChIP-seq data set, processed as described above (Lambing et al., 2019), an MNase-seq and single-end SPO11-1-oligonucleotide libraries processed as previously described (Choi et al., 2018), and a set of 3,320 crossovers derived from genotyping-by-sequencing of recombinant Col-0 × L*er*-0 F_2 populations (Serra et al., 2018a; Underwood et al., 2018). For comparisons with histone modifications, H3K9me2, H3K27me1, H3K27me3 libraries were generated by Dr Christophe Lambing and kindly made available (Lambing et al., 2019). Previously published H3K4me3 ChIP-seq and H2A.W ChIP-seq datasets were also utilised (Yelagandula et al., 2014; Choi et al., 2018). For expression analysis of gene hexiles, based on HA-MSH4 enrichment, a previously published RNA-seq dataset from floral bud tissues was used (Choi et al., 2018). For DNA methylation analysis, single-nucleotide-resolution DNA methylation proportions were calculated from published bisulfite sequencing libraries (Stroud et al., 2013).

To evaluate the overlap between HA-MSH4 ChIP-seq peaks and SPO11-1-oligo hotspots at transposable element superfamilies, permutation tests were performed using the R package regioneR version 1.6.2 (Gel et al., 2016; Choi et al., 2018). The number of overlaps were defined by the frequency of HA-MSH4 or SPO11-1-oligo hotspots overlapping with one or more

transposable elements of the specified superfamily. For the permutation tests, 10,000 sets of random genomic loci with the same width distribution as either the HA-MSH4 peaks or SPO11-1-oligo hotspots were defined. Thus, the average number of random loci overlapping the transposon superfamily (the expected) was compared to the number of overlapping peaks or hotspots (the observed). *P*-values denoting the significance of the observed overlap for the permutation tests were $\geq 1\times10^{-4}$ or higher, because the peak loci are randomised (permuted) 10,000 times.

3.4.3 Basic statistical tests and data visualisation

All basic statistical tests and data visualisation were performed with R (version 3.4.1).

Chapter Three – Results – Analysis of meiotic crossover frequency within multiple genetic intervals in the Arabidopsis *msh2-1* mutant

3.1 Introduction

In plants, DNA repair processes are essential for maintaining genome stability (Puchta & Hohn, 2012). MMR has been demonstrated to protect against spontaneous mutations and hence increases the fidelity of genetic inheritance (Leonard et al., 2003; Belfield et al., 2018). Crucial to this process are the MutS-homologue (MSH) family that function to recognise mismatched nucleotides that are mis-incorporated during DNA replication (Kunkel & Erie, 2015). In Arabidopsis, there are four MSH homologs involved in MMR – MSH2, MSH3, MSH6 and MSH7 – which form three heterodimers, all with MSH2 as a common subunit (Culligan & Hays, 2000; Hays, 2002). These heterodimers show differing binding affinity for certain types of mismatch, but primarily recognise single base pair mismatches or small indels (Wu et al., 2003). Consistent with other eukaryotic model systems, mutation rates are elevated in the Arabidopsis *msh2-1* loss-of-function mutant (Kunkel & Erie, 2015; Watson et al., 2016; Belfield et al., 2018). More specifically, the mutational spectra in *msh2-1* is similar to that of the wild type, with GC:AT transitions dominating (Watson et al., 2016; Belfield et al., 2018). However, the occurrence of point mutations increases by ~170-fold in the *msh2-1* background, whereas that of indels increases by >1000-fold and is predominated by single nucleotide deletions (Belfield et al., 2018). Like other organisms, MMR acts to repress somatic homologous recombination between divergent sequence repeats in Arabidopsis (Emmanuel et al., 2006; Li et al., 2006). In addition, MMR has been shown to antagonise meiotic crossover recombination (Emmanuel et al., 2006). For example, a previous study of meiotic crossover frequency in the Arabidopsis *msh2-1* mutant revealed that meiotic crossovers increase by 40% in a 2 Mb sub-telomeric genetic interval, in a Col × L*er* F_1 hybrid (Emmanuel et al., 2006). Although providing an exciting insight, this study was limited to the analysis of a single genetic interval and did not investigate the role of *msh2-1* in an inbred background.

Therefore, I sought to extend investigation of the role of MMR in meiotic recombination in Arabidopsis in several ways. Firstly, to confirm the increased crossover frequency in *msh2-1* hybrids at several additional genetic intervals located in the chromosome arms. Secondly, I sought to extend this analysis to investigate additional chromosome regions, such as the pericentromere, which are known to be refractory to genome-wide increases in crossover frequency in other anti-crossover mutants (Fernandes et al., 2018a; Serra et al., 2018a). Thirdly, I sought to investigate the effects of *msh2-1* on meiotic recombination in several different Arabidopsis hybrids, which differed in levels of polymorphism and genetic structure, and compare these to inbreds. Finally, I sought to screen for cytological or fertility phenotypes

in *msh2-1* hybrids, to identify any defects due to loss of MMR activity. Together, the aim of these investigations was to provide a better understanding of the relationship between sequence diversity, meiotic crossover recombination, and MMR.

3.2 Three Arabidopsis accessions were selected to generate experimental hybrids

Three Arabidopsis accessions were selected, which would produce differing patterns of interhomolog polymorphism when crossed with the Col-0 accession (fig. 12A). The L*er*-0 accession, which was collected in Poland, was chosen due to its widespread use in meiotic recombination experiments (e.g. Fernandes et al., 2018a; Serra et al., 2018a), its high quality genome assembly and polymorphism maps (Zapata et al., 2016), and its previous use to investigate the role of MSH2 in regulating crossover frequency (Emmanuel et al., 2006). The Ct-1 accession was chosen as previous mapping experiments indicated that no natural *trans* acting modifiers of recombination frequency existed in this genetic background (Ziolkowski et al 2015), and because the accession was being utilised for relevant experiments in other laboratories (Dr Piotr Ziolkowski, personal communication). Furthermore, realignment of DNA sequencing data from the Ct-1 and L*er*-0 accessions indicated a similar *MSH2* copy number to Col-0 (5.38, 4.26, and 5.32 RPKM, respectively), suggesting that additional copies of *MSH2* did not exist in either accession, which might otherwise confound the introgression approach (Alonso-Blanco et al., 2016). I was prompted to address this concern due to a report indicating extensive copy number variation in a region spanning the *MSH2* locus (Zmienko et al., 2016).

Finally, I used a third more complex genetic background, in order to investigate the effects of *trans* modifiers of crossover frequency (Lawrence et al., 2017). This line was derived from the Cvi-0 accession, and was previously reported as a pure Cvi-0 inbred (Ziolkowski et al., 2015). However, re-sequencing of this line revealed it to contain multiple L*er*-0 introgressions and two Col-0 introgressions (fig 12B). The entirety of chromosome 5 is substituted for L*er*-0, and small regions of L*er*-0 sequence existed on chromosomes 1 (0 – 3.7 and 26.5 – 30.4 Mb), 2 (0.7 – 7.2 Mb) and 4 (0 – 1.6 Mb). In addition, this line contains a small Col-0 introgression on chromosomes 2 (7.2 – 8.0 Mb) and a larger Col-0 introgression on chromosome 4 (12.5 – 18.5 Mb). Hereafter, I termed this line CLC (Cvi-0, L*er*-0, Col-0) (fig. 12B). The boundaries of genotype change across the CLC genome were determined by assessing the correspondence between Cvi-0 or L*er*-0 SNP coordinates and CLC SNP coordinates, and identifying regions with altered correspondence. Regions of Col-0 were identified due to the absence of SNPs.

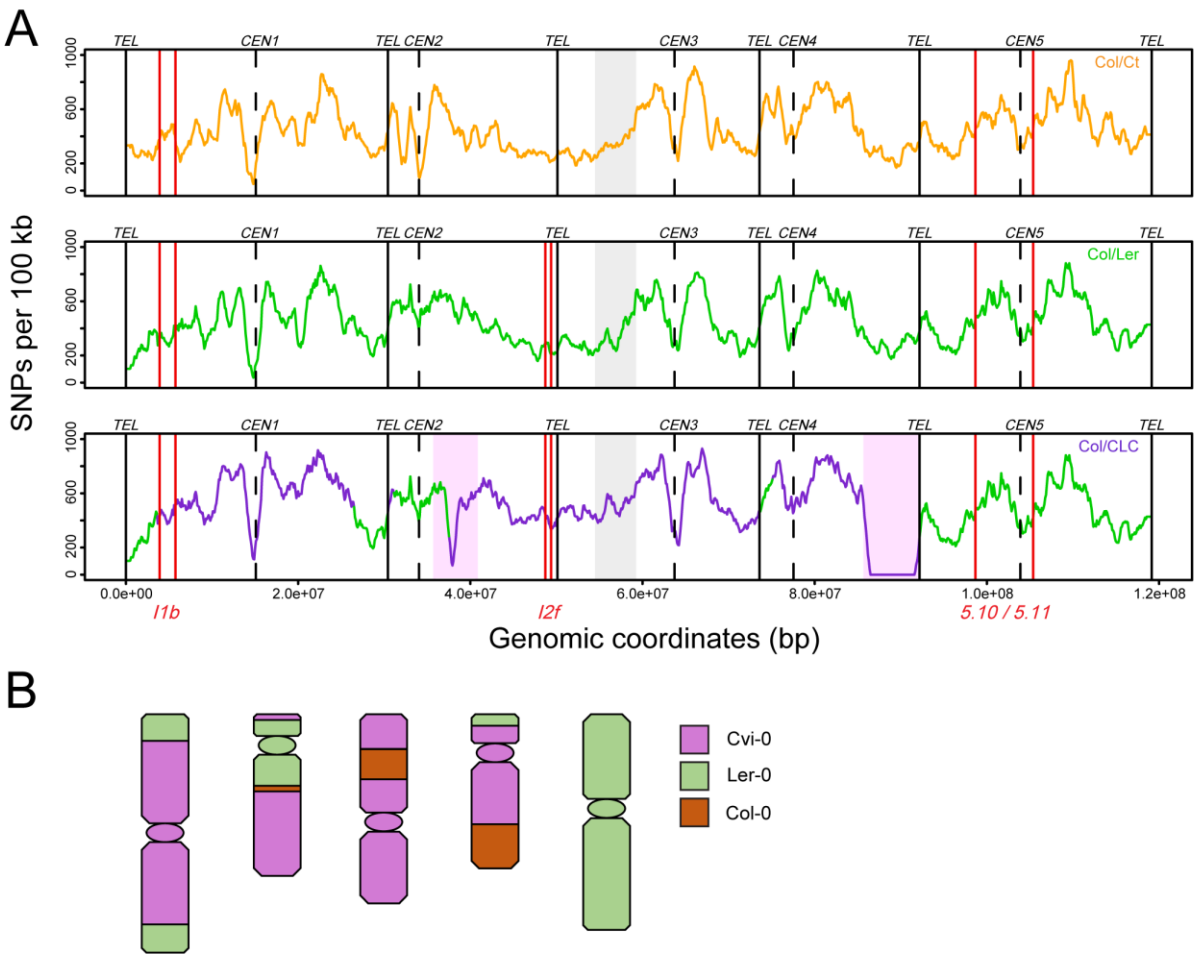


Figure 12. Diagram of the experimental system and CLC karyotype.

(A) A schematic representation of the experimental system, showing SNP density maps for the three different Arabidopsis hybrids and the location of three genetic intervals used for crossover analysis. SNP densities relative to Col-0 are shown in 100 kb sliding windows, for Ct-1 (orange), L*er*-0 (green) and CLC (Cvi-0: purple; L*er*-0: green) based on the 1,135 genomes datasets (Alonso-Blanco et al., 2016). Telomeres (vertical black lines), centromeres (dashed black lines), FTL/CTL genetic intervals (red lines) (Francis et al., 2007; Wu et al., 2015), the region of *msh2-1* T-DNA Col-0 introgression (grey boxes), and the additional Col-0 introgressed regions of CLC (pink boxes) are indicated. (B) A schematic representation of the CLC karyotype.

These three genetic backgrounds have differing genome-wide SNP distributions, whose key parameters are summarised in table 2. Genome-wide sequence divergence at the level of SNPs is similar for all three backgrounds, at 0.45, 0.45 and 0.5 % in Ct-1, L*er*-0 and CLC, respectively. However, at the chromosome scale there is considerable variability in SNP

density (fig. 12A). SNP density is consistently high in the centromere proximal regions, as reported (Alonso-Blanco et al., 2016). It is also notable that the Arabidopsis centromeres remain unassembled, and consist of repetitive Mb arrays of *CEN180* satellite repeats (AGI, 2000; Hosouchi et al., 2002). Hence, the drop in SNP density over these regions may be an artefact due to the challenges of mapping short read sequencing data to these regions, and indeed structural variation in centromere satellite arrays is observed between accessions (Ito et al., 2007). Regional differences in polymorphism density are seen between the three accessions (fig. 12A). Notably, the CLC genome is predominantly (81%) derived from Cvi-0, an Arabidopsis 'relict' accession collected on Cape Verde (Alonso-Blanco et al., 2016). Relict Arabidopsis accessions are found in locations where environmental conditions remained more constant during the last glacial period, such as the Iberian Peninsula, and it is thought that extant Arabidopsis accessions are derived from expansions from these glacial refugia (Alonso-Blanco et al. 2016). Along with a second island accession (Can-0), Cvi-0 was found to have a genome diverged from all other accessions analysed (Nordborg et al., 2005; Alonso-Blanco et al., 2016). Hence, its level of divergence relative to Ct-1 and L*er*-0 is likely underestimated from SNPs and indels called using short read sequencing analyses, given their limited ability to identify larger scale structural polymorphism, such as inversions, translocations, large scale deletions and copy number variants (Alonso-Blanco et al., 2016; Zapata et al., 2016). This greater divergence was reflected when comparing indel frequencies between CLC, Ct-1 and L*er*-0: the CLC genome contained ~30% more indels than Ct-1 or L*er*-0 relative to Col-0 (table 2).

SNPs										
	Genome		***I1b***		***I2f***		***5.10***		***5.11***	
Accession	**Total**	**Per kb**	**Total**	**Per kb**	**Total**	**Per kb**	**Total**	**Per kb**	**Total**	**Per kb**
Ct-1	532,319	4.47	8,275	9.73	1,718	NA	33,831	5.03	34,695	4.98
L*er*-0	530,322	4.45	5,884	6.92	2,028	3.03	34,880	5.18	35,282	5.06
CLC	599,338	5.03	7,924	9.32	2,437	3.64	34,880	5.18	35,282	5.06
Indels										
Ct-1	38,870	0.33	733	0.86	136	NA	2,090	0.31	2,119	0.30
L*er*-0	38,419	0.32	482	0.57	191	0.28	2,086	0.31	2,110	0.30
CLC	51,037	0.43	808	0.95	259	0.39	2,086	0.31	2,110	0.30
	SNP intervals									
	Mean	**Median**	**Max**	**StDev**						
Ct-1	224	53	158,982	918						
L*er*-0	225	54	197,227	962						
CLC	187	51	131,982	719						

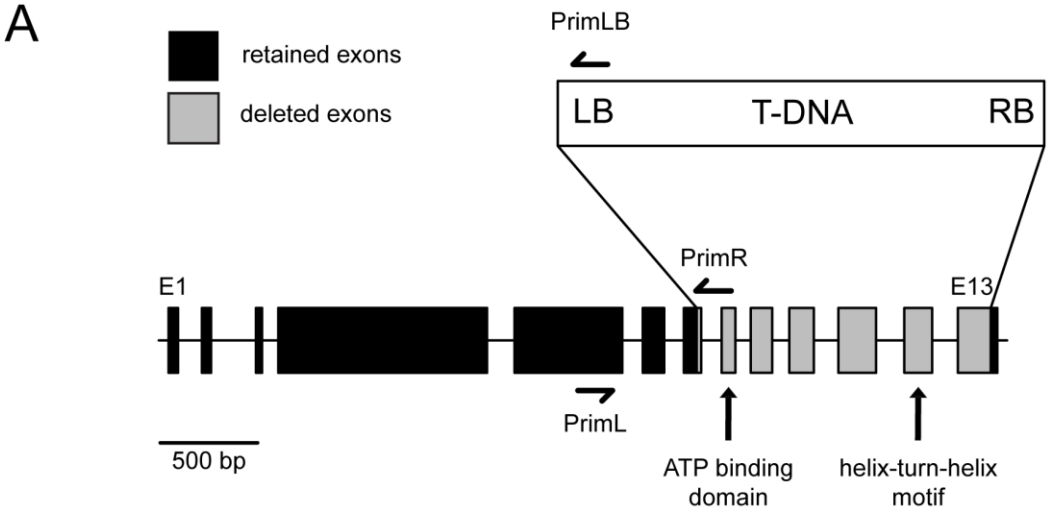
Table 2. Description of polymorphism in Ct-1, L*er*-0 and CLC accessions relative to Col-0.

Total SNP and indel numbers relative to Col-0 are provided for the Ct-1, L*er*-0 and CLC genomes, and the *I1b*, *I2f*, *5.10* and *5.11* genetic intervals. Density of SNPs/kb is also provided. The metrics of mean, median, maximum and standard deviation (in base pairs) are also provided for the distances between adjacent SNPs ('SNP intervals'). These values were calculated from the 1,135 genomes datasets (Alonso-Blanco et al., 2016).

3.3 Introgression of the *msh2-1* mutation into three divergent Arabidopsis accessions

I introgressed the *msh2-1* mutation into these different genetic backgrounds, using a backcrossing strategy (Leonard et al., 2003; Watson et al., 2016; Belfield et al., 2018). I selected a *msh2* T-DNA insertion mutant (*msh2-1*) in the Col-0 accession, previously characterised as a loss-of-function allele causing an elevated spontaneous mutation rate and microsatellite instability (Leonard et al., 2003; Watson et al., 2016; Belfield et al., 2018). The *msh2-1* T-DNA insertion interrupts the *MSH2* coding sequence, and causes a deletion of exons 8 to 12 and parts of exons 7 and 13 (fig. 13A) (Leonard et al., 2003). Col-0 plants heterozygous for the *msh2-1* mutation (*msh2-1*/+) were crossed as a male parent to one of the three wild type parental accessions (Ct-1, L*er*-0 or CLC). The resulting F_1 progeny were then backcrossed (BC) as a male parent to their parental accessions for a further 5 generations (fig. 13B). With the exception of two rounds of mitosis in the male gametophytic generation, the *msh2-1* mutation was maintained as a heterozygote throughout this process.

To confirm the *msh2-1* introgression, simple sequence length polymorphism (SSLP) and cleaved amplified polymorphic sequences (CAPS) PCR genotyping was performed using a set of markers distributed throughout the genome (fig. 14) (appendix 7.1). The SSLP approach amplifies short sequences known to contain indel polymorphisms of 25-100 bp between the two accessions of interest. Hence, assessing PCR product sizes via gel electrophoresis enables the assignment of an AA homozygous, AB heterozygous, or BB homozygous genotype at the marker position. The CAPS approach is analogous, but relies on the presence of restriction site polymorphisms within the amplicon, cleavage of which results in an altered product size. This analysis confirmed that, with the exception of a ~5 Mb Col-0 region surrounding the *msh2-1* introgression, the genome was fixed for the parental accession at all markers tested. In addition, I PCR amplified the T-DNA insertion site using primers targeting the T-DNA left border and a genomic sequence internal to the endogenous *MSH2* gene, from genomic DNA extracted from the BC introgression lines. This PCR product was Sanger sequenced and the sequencing trace confirmed the presence of the *msh2-1* T-DNA.


A
retained exons
deleted exons
PrimLB
LB
T-DNA
RB
PrimR
E1
E13
PrimL
500 bp
ATP binding domain
helix-turn-helix motif

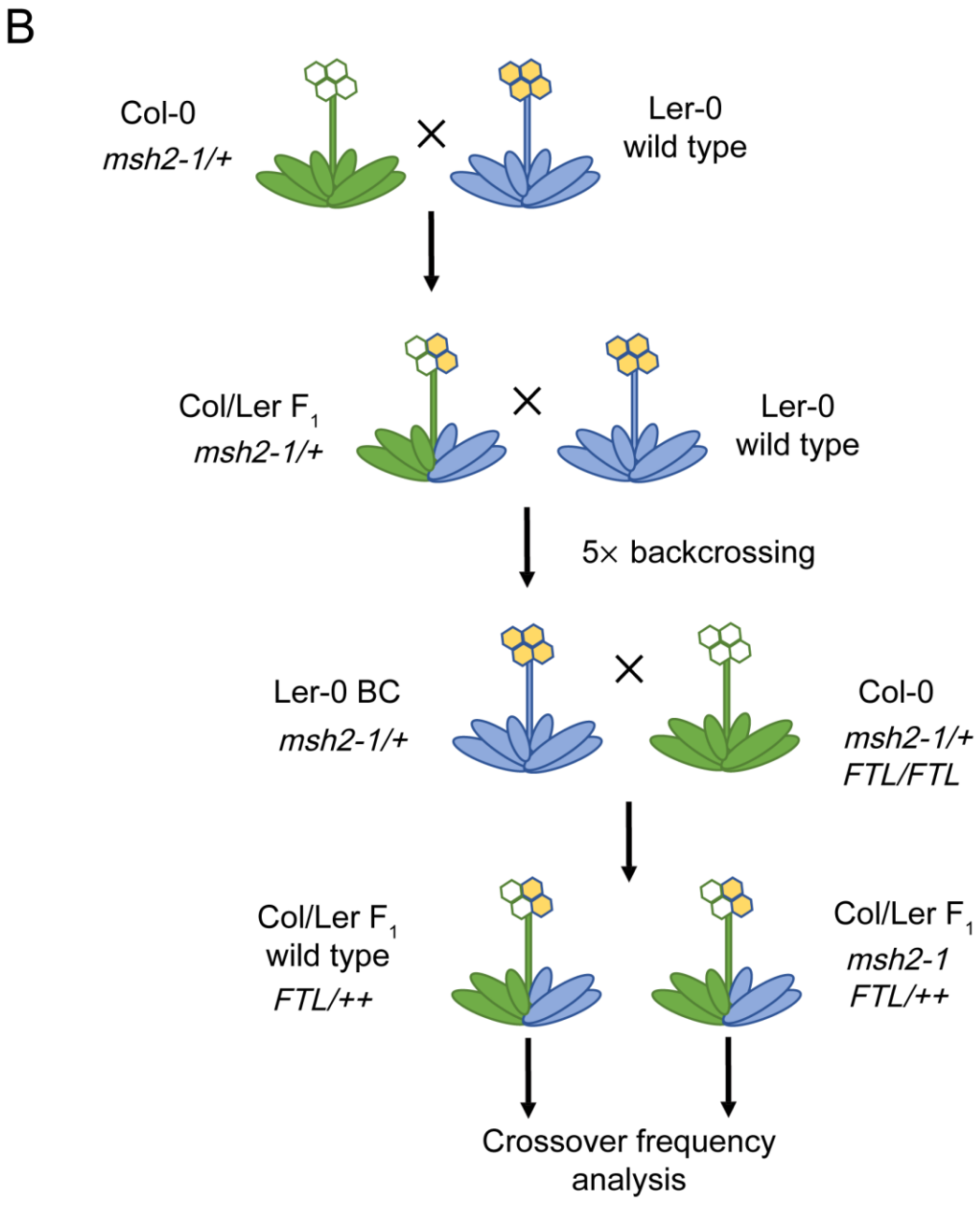

B
Col-0
msh2-1/+
Ler-0
wild type
Col/Ler F1
msh2-1/+
Ler-0
wild type
5× backcrossing
Ler-0 BC
msh2-1/+
Col-0
msh2-1/+
FTL/FTL
Col/Ler F1
wild type
FTL/++
Col/Ler F1
msh2-1
FTL/++
Crossover frequency analysis

Figure 13. Diagram of the Col-0 *msh2-1* T-DNA and introgression strategy.

(A) A schematic representation of the *MSH2* locus from exon 1 (E1) to exon 13 (E13). The T-DNA insertion begins partway through exon 7 and has deleted the following 1,510 bp of sequence, including the helix-turn-helix and ATP binding domains essential for MMR activity (Alani et al., 1997; Leonard et al., 2003). Retained exons are shown in black and deleted exons in grey. Noncoding sequence is represented by straight black lines. The orientation of the T-DNA's left border (LB) and right border (RB) is also indicated. Primers used for genotyping the wild type product ('PrimR' + 'PrimL') and the T-DNA product ('PrimL' + 'PrimLB') are indicated. (B) Diagram of backcrossing strategy used to introgress *msh2-1* T-DNA from the Col-0 background (green and white) into either the Ct-1, L*er*-0 or CLC accessions (blue and yellow). The L*er*-0 accession is chosen as a representative. Col/Ler F_1 hybrids were recurrently backcrossed (BC) to a stock L*er*-0 accession for five generations, selecting for the *msh2-1* T-DNA in each generation by genotyping. L*er*-0 BC *msh2*-1/+ plants were then crossed to *msh2-1*/+ Col-0 plants homozygous for a selected FTL or CTL interval. This produced 'F_1' hybrids either wild type or MMR-deficient, in both cases hemizygous for the FTL or CTL reporters, which could be scored to crossover frequency.

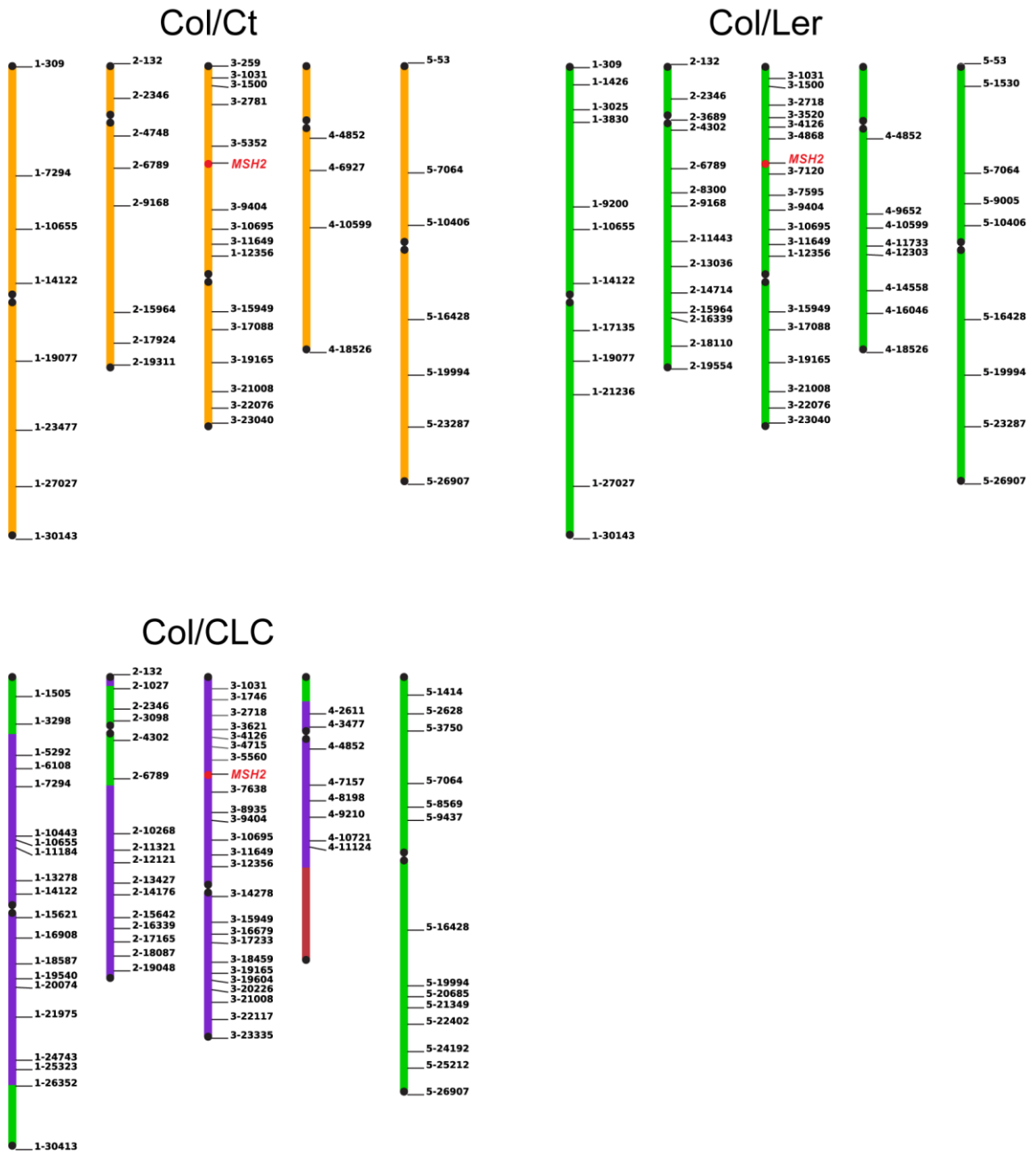


Figure 14. Schematic of the five Arabidopsis chromosomes, showing the position of genetic markers used to monitor introgression of the *msh2-1* T-DNA.

A combination of SSLP and CAPS markers were used to confirm a homozygous Ct-0 (orange), L*er*-0 (green) or CLC (purple/green) genotype. A region of Col-0 sequence in the CLC background is shown in dark red. The positions of genetic markers are indicated. A red dot indicates the position of the *MSH2* genomic locus.

3.4 Crossover frequency is increased in *msh2-1* hybrids at two sub-telomeric intervals

I first sought to reproduce the crossover increase previously reported in a Col-0 × L*er*-0 F_1 hybrid *msh2-1* mutant (Emmanuel et al., 2006). In order to measure crossover frequency, I utilised the pollen-based fluorescent reporter FTL intervals *I1b*, *I2f and I5a*. *I1b* is a 1.85 Mb interstitial interval on chromosome 1, *I2f* is a 0.67 Mb sub-telomeric interval on chromosome 2, and *I5a* is a 4.92 Mb sub-telomeric interval on chromosome 5, respectively (fig. 12A). These intervals provide a powerful approach for measuring meiotic crossover frequency (Francis et al., 2007; Yelina et al., 2013). Each interval is defined by T-DNA insertions containing a gene expressing either YFP or RFP from the *LAT52* promoter, which causes specific fluorescence in pollen (Francis et al., 2007). When hemizygous (i.e. present on one chromosome and not the other) and in *cis* configuration (i.e. both linked on the same chromosome) a crossover within the interval causes the segregation of T-DNAs into different gametes and hence the generation of single colour pollen grains (Francis et al., 2007). Fluorescence can also be assayed using flow cytometry, enabling high throughput measures of crossover frequency (Yelina et al., 2013). In comparison to seed-based CTL measurements of crossover frequency, the FTL-flow-cytometry approach enables deeper, and therefore more powerful, measurements of crossover frequency. However, one disadvantage is that they only allow crossover frequency to be measured in the male meiocytes, and crossover frequency and distribution are known to vary between sexes in Arabidopsis (Giraut et al., 2011; Fernandes et al., 2018a).

To generate experimental hybrids, the introgressed *msh2-1*/+ (L*er*-0) plants were crossed to parental *msh2-1*/+ (Col-0) lines which were also homozygous for a reporter interval (fig. 13B). The progeny were genotyped for the *msh2-1* T-DNA and phenotyped for pollen fluorescence, in order to identify plants that were hemizygous for the fluorescent transgenes and hence scorable for crossover frequency. Pollen was collected from scorable wild type and *msh2-1* mutant plants, and filtered to remove any somatic tissue fragments. Flow cytometry was performed to measure the frequency of recombinant (R) and non-recombinant classes (NR) – i.e. double-colour (NR), single-colour red (R), single colour yellow (R), non-colour (NR) – and these values were used to calculate crossover frequency (see section 2.1.5).

I first tested whether *msh2-1* influenced crossover frequency in an otherwise inbred Col-0 background, at the *I1b*, *I2f* and *I5a* intervals. This revealed no significant difference between the *msh2-1* and wild type genotypes at the *I1b*, *I2f* and *I5a* intervals (P = 0.071, 0.548 and 0.534 respectively; Mann-Whitney U tests) (fig. 15) (appendix 7.3). This confirmed the prediction that loss of MMR activity would have no impact on crossover frequency in an inbred background.

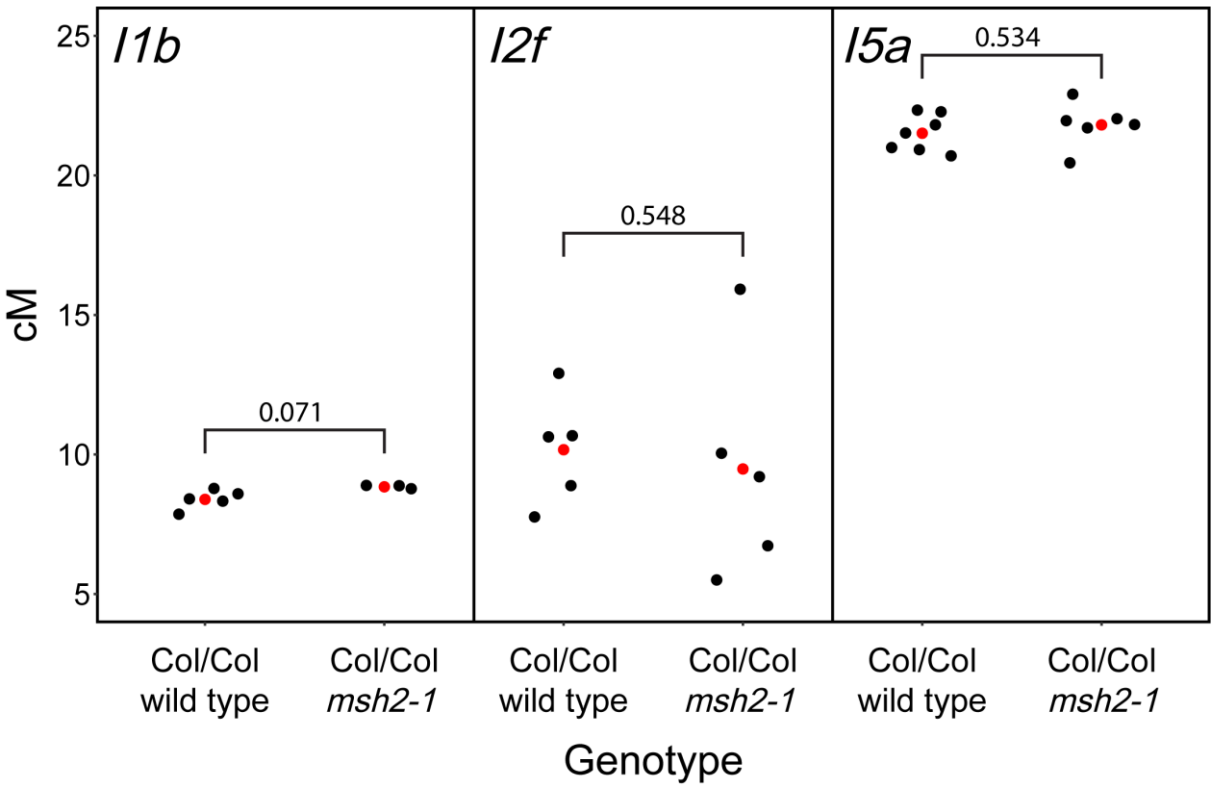


Figure 15. The *msh2-1* mutation has little to no effect in an inbred Col-0 background.

The genetic distance in centimorgans (cM) at *I1b*, *I2f* and *I5a* was measured in wild type and *msh2-1* Col-0 inbreds. Black dots represent replicate measurements and red dots represent the mean of each genotype. Mann-Whitney U tests were performed to test for significant differences between genotypes.

To test the impact of polymorphism on crossover frequency, I then analysed crossover frequency at the genetic intervals *I1b* and *I2f* in a Col/Ler hybrid background, comparing wild type and *msh2-1*. I observed that the *msh2-1* mutation caused a 26.1% increase in crossover frequency at *I1b* (from 5.44 to 6.85 cM) and a 29.1% increase at *I2f* (from 6.66 to 8.60 cM) in a Col/Ler hybrid, compared to the wild type (P = 0.0005 and 0.0057, respectively; Mann-Whitney U tests) (fig. 16, 17) (appendix 7.4 & 7.5). These results are in broad agreement with the 40% increase in crossover frequency previously observed in the arm of chromosome 5 of a Col/Ler *msh2-1* hybrid (Emmanuel et al., 2006), and are consistent with MSH2 suppressing crossovers specifically in a hybrid genetic context.

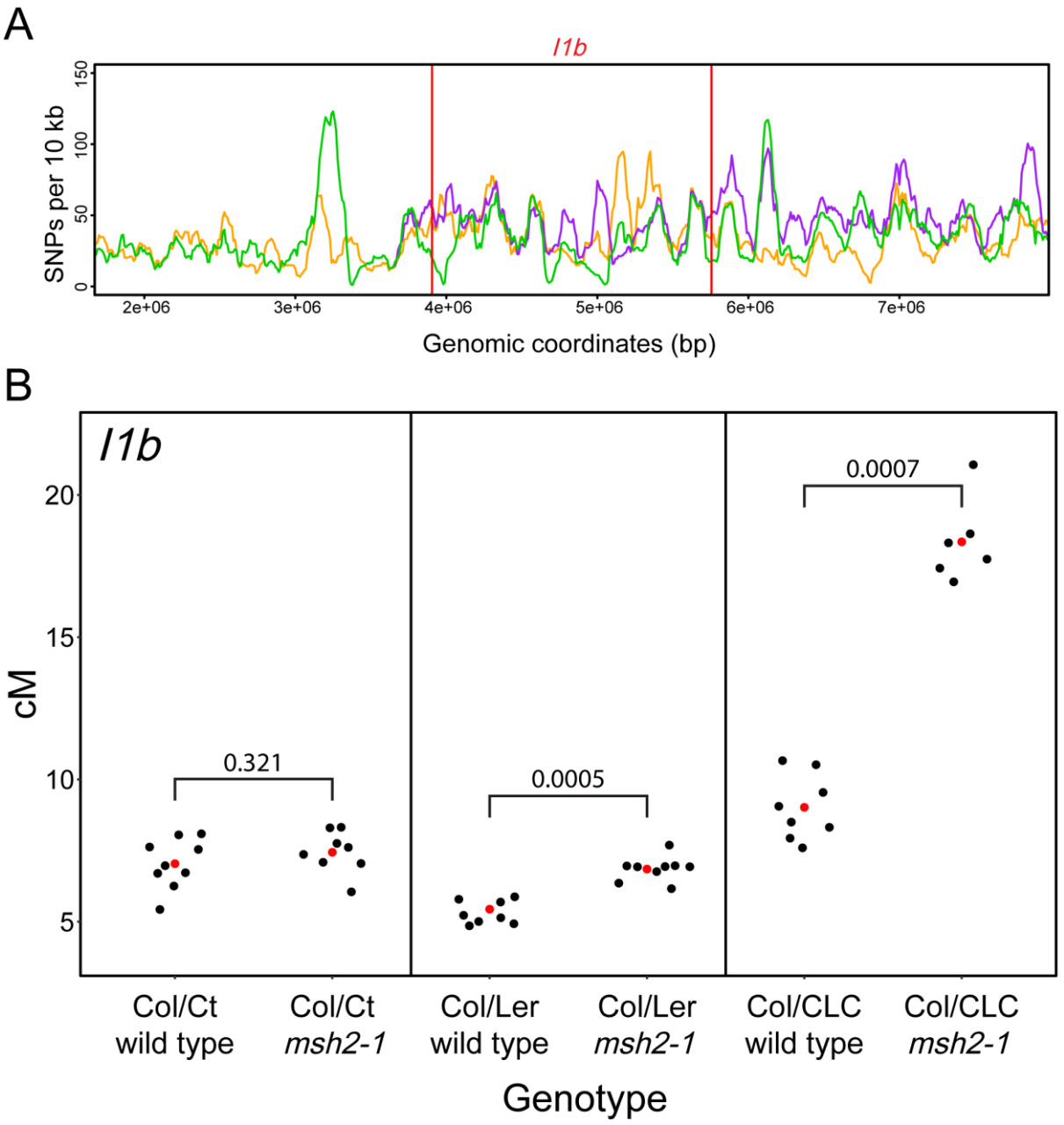


Figure 16. The *msh2-1* mutation increases meiotic crossover frequency at *I1b* in a hybrid context.

(A) SNP density is plotted for three Arabidopsis hybrids across the *I1b* genetic interval used for crossover analysis (red lines), and immediately adjacent regions. SNP densities relative to Col-0 are shown in 10 kb sliding windows, for Ct-1 (orange), L*er*-0 (green) and CLC (Cvi-0: purple; L*er*-0: green) based on the 1,135 genomes datasets (Alonso-Blanco et al., 2016). (B) The genetic distance in centimorgans (cM) at *I1b* was measured in Col/Ct, Col/Ler and Col/CLC wild type and *msh2-1* mutants. Black dots represent replicate measurements and red dots represent the mean of each genotype. Mann-Whitney U tests were performed to test for significant differences between genotypes.

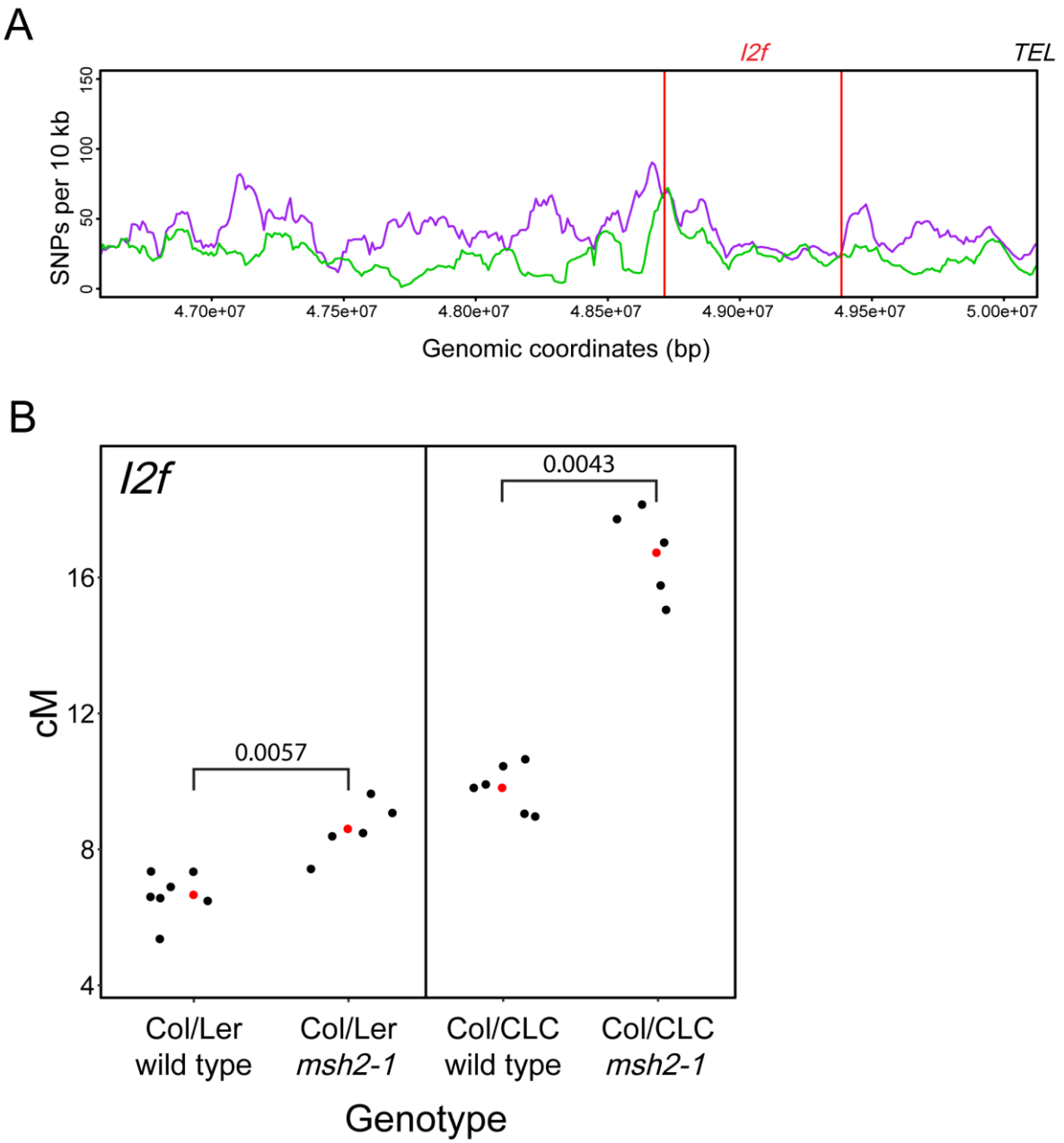


Figure 17. The *msh2-1* mutation increases meiotic crossover frequency at *I2f* in a hybrid context.

(A) SNP density is plotted for two Arabidopsis hybrids across the *I2f* genetic interval used for crossover analysis (red lines), and immediately adjacent regions. SNP densities relative to Col-0 are shown in 10 kb sliding windows, for L*er*-0 (green) and CLC (Cvi-0: purple; L*er*-0: green) based on the 1,135 genomes datasets (Alonso-Blanco et al., 2016). (B) Genetic distance was measured at *I2f* in Col/Ler and Col/CLC wild type and *msh2-1* mutants. Black dots represent replicate measurements and red dots represent the mean of each genotype. Mann-Whitney U tests were performed to test for significant differences between genotypes.

I repeated this approach using two additional hybrids, Col/Ct and Col/CLC. At *I1b*, crossover frequency was slightly increased by 5.7% in the Col/Ct *msh2-1* mutant, relative to the wild type (from 7.04 to 7.44 cM), however this increase was not statistically significant ($P = 0.321$; Mann-Whitney *U* test) (fig 16B). The smaller, non-significant increase in the Col/Ct *msh2-1* background relative to the Col/Ler hybrid was surprising, given that Col/Ct has ~40% more interhomolog polymorphism (when counting both SNPs and indels) within the *I1b* interval relative to Col/Ler. Unfortunately, it was not possible to test the effects of *msh2-1* on crossover frequency at *I2f* in Col/Ct, due to insufficient seed production when attempting to generate this hybrid line. However, crossover frequency was dramatically increased in the Col/CLC *msh2-1* mutant, increasing by 70.6% at *I1b* and 103% at *I2f* ($P = 0.0007$ and 0.0043, respectively; Mann-Whitney *U* tests) (fig. 16, 17) (appendix 7.4 & 7.5). Furthermore, crossover frequency in the wild type was significantly elevated in this hybrid, relative to Col/Ler and Col/Ct, consistent with previous reports (Ziolkowski et al., 2015).

These observations are broadly consistent with MMR acting to antagonise meiotic recombination in a hybrid context, with limited effect in an inbred context. However, there is significant variation in *msh2-1* dependency between intervals and hybrid accessions. From this analysis, the relationship with polymorphism density and the relative increases in *msh2-1* crossover frequency are unclear, particularly given that the hybrid with the greatest interhomolog polymorphism at *I1b* (Col/Ct) showed no significant increase in crossover frequency in the *msh2-1* mutant. However, it is possible that additional, unknown factors are limiting crossover levels in the Col/Ct hybrid epistatic to the activity of MSH2, or that the current analysis did not have sufficient statistical power to detect a significant difference. The most dramatic changes in the *msh2-1* mutant were seen for the Col/CLC hybrid, perhaps suggesting that small indels, known to be a target for MMR system (Wu et al., 2003), play a greater role in determining crossover locations than SNPs.

3.5 Crossover frequency is decreased in *msh2-1* hybrids at two pericentromeric intervals

Given that crossover frequency increased in *msh2-1* at two FTL intervals located in the gene-rich chromosome arms, I predicted that crossover frequency would increase to a greater extent in the *msh2-1* hybrids in the pericentromeric regions, as these regions have a higher polymorphism level (fig. 12A) (Simoens et al., 1988; AGI, 2000; Hosouchi et al., 2002; Ito et al., 2007). To test this, I utilised the crossover reporters *5.10* and *5.11*, which are large seed-based CTL intervals spanning the pericentromere of chromosome 5, and are 6.7 and 7.0 Mb, respectively. As 98% of the *5.11* interval overlaps *5.10*, this interval served as an independent replicate. The CTL system is analogous to the FTL system, but fluorescence is measured in seeds rather than pollen, driven from the *NapA* promoter, providing an average crossover

frequency for male and female meiosis when F_2 seed is analysed (Melamed-Bessudo et al. 2005; Berchowitz & Copenhaver, 2008; Wu et al., 2015). Although SNP density is not markedly higher at *5.10* and *5.11* compared to *I1b* and *I2f* (table 2), as calculated from the 1,135 genomes data (Alonso-Blanco et al., 2016), this is thought to be a significant underestimation of the diversity in the Arabidopsis centromeres and pericentromeres, given the challenges of mapping and assembling these regions (see section 3.7 for a more detailed discussion). I derived hybrids between Col-0 *msh2-1*/+ plants that were hemizygous for either the *5.10* or *5.11* CTL transgenes and L*er*-0, CLC, or Ct-1 backcrossed *msh2-1*/+ plants, in a crossing strategy analogous to that described above (fig. 12B). Hybrid F_1 seed was microscopically screened to select for hemizygous reporter transgenes, and plants from this generation were genotyped. Seed was collected from wild type and *msh2-1* mutants, and individuals were scored for their crossover frequency (see section 2.1.6).

Surprisingly, crossover frequency at the *5.10* interval was reduced by 8.6%, 4.5%, and 8.5% in the *msh2-1* mutant Col/Ler, Col/CLC, and Col/Ct hybrids, respectively (fig. 18B) (appendix 7.6). These reductions were statistically significant in all cases (P = 0.014, 0.045, and 0.036, respectively; Mann-Whitney U tests). In contrast, I observed no significant decrease in crossover frequency in the Col-0 inbred *msh2-1* mutants, relative to the wild type (P = 0.492; Mann-Whitney U test) (fig. 18B) (appendix 7.6). These results were broadly recapitulated at the *5.11* interval (fig. 18C) (appendix 7.7), which is expected as these CTL intervals are overlapping. Crossover frequency at *5.11* was significantly reduced by 7.2% and 13.4% in the *msh2-1* mutant Col/Ler and Col/CLC hybrids, respectively (P = 0.018 and 0.036, respectively; Mann-Whitney U tests) (fig. 18C), whilst remaining unchanged in the Col-0 inbred background (P = 0.490; Mann-Whitney U test). However, the 7.4% decrease observed for the Col/Ct population was not statistically significant (P = 0.370; Mann-Whitney U test), although the direction of change was consistent with the other hybrid *msh2-1* mutants tested. One likely explanation for the lack of a statistically significant change is that the Col/Ct hybrid data appeared particularly variable for the *5.11* interval, perhaps due to an unknown environmental factor influencing seed quality or disturbing meiosis. Taken together, this analysis of two overlapping pericentromeric intervals suggests a more complex relationship between polymorphism and crossover frequency in the *msh2-1* mutant, where its effects are region and/or SNP density specific, and indicate that crossovers may be redistributing across the genome. Furthermore, these reductions are even more surprising given the high sequence divergence around the Arabidopsis pericentromere, as I predicted more divergent genomic regions to show the greatest increase in crossover frequency in *msh2-1*.

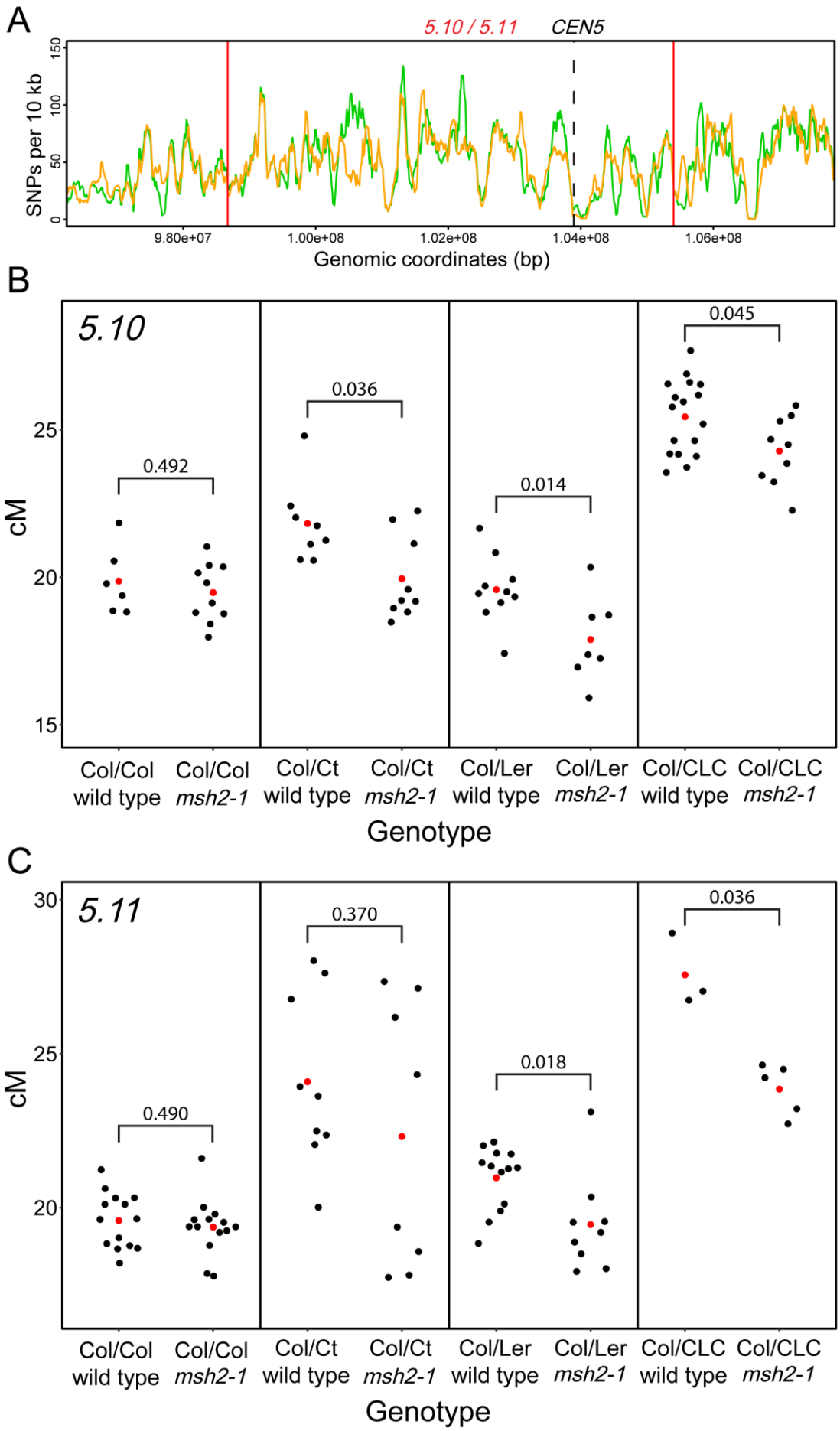

A
5.10 / 5.11
CEN5
SNPs per 10 kb
0
50
100
150
9.80e+07
1.00e+08
1.02e+08
1.04e+08
1.06e+08
Genomic coordinates (bp)
B
5.10
cM
15
20
25
0.492
0.036
0.014
0.045
Col/Col wild type
Col/Col msh2-1
Col/Ct wild type
Col/Ct msh2-1
Col/Ler wild type
Col/Ler msh2-1
Col/CLC wild type
Col/CLC msh2-1
Genotype
C
5.11
cM
20
25
30
0.490
0.370
0.018
0.036
Col/Col wild type
Col/Col msh2-1
Col/Ct wild type
Col/Ct msh2-1
Col/Ler wild type
Col/Ler msh2-1
Col/CLC wild type
Col/CLC msh2-1
Genotype

Figure 18. The *msh2-1* mutation decreases crossover frequency at two centromeric genetic intervals.

(A) SNP density is plotted for three Arabidopsis hybrids across the *5.10/5.11* genetic intervals used for crossover analysis (red lines), and immediately adjacent regions. SNP densities relative to Col-0 are shown in 10 kb sliding windows, for L*er*-0 (green) and CLC (Cvi-0: purple; L*er*-0: green) based on the 1,135 genomes datasets (Alonso-Blanco et al., 2016). For these intervals, the SNP density is identical for L*er*-0 and CLC. The location of the centromere is shown with dashed lines. (B) The genetic distance in centimorgans (cM) at *5.10* was measured in wild type and *msh2-1* Col-0 inbreds and Col/Ct, Col/Ler and Col/CLC hybrids. (C) The genetic distance in centimorgans (cM) at *5.11* was measured in wild type and *msh2-1* Col-0 inbreds and Col/Ct, Col/Ler and Col/CLC hybrids. Black dots represent replicate measurements and red dots represent the mean of each genotype. Mann-Whitney U tests were performed to test for significant differences between genotypes.

3.6 Pollen viability is reduced in the inbred and hybrid *msh2-1* backgrounds

The changes in crossover frequency observed in *msh2-1* prompted me to address effects on fertility. To measure the fertility of male gametes, I performed Alexander staining on pollen grains collected from nine plants per genotype. For each sample, three mature flowers were selected from the primary floral axis, and pollen was pooled during collection. Viable pollen grains stain purple, whereas inviable pollen tends to be misshapen and stains blue-green (Alexander, 1969). A mean of 1,865 pollen grains were scored per sample (appendix 7.8). Small, but significant, increases in pollen inviability were observed in the *msh2-1* mutant for all inbred backgrounds tested, by 1.13%, 3.03% and 2.45%, in Col-0, L*er*-0 BC, and CLC BC, respectively ($P = 5.7\times10^{-11}$, $< 2.2\times10^{-16}$, and $< 2.2\times10^{-16}$, respectively; X^2 tests) (fig. 19). For the Col/CLC hybrids, the *msh2-1* mutation significantly increased inviability by 0.82%, compared to the wild type ($P = 5.18\times10^{-8}$; X^2 test) (fig. 19). Together, these results indicated that loss of viability was not dependent on genome-wide heterozygosity and therefore unlikely to be related to the crossover phenotypes observed.

3.7 Cytological characterisation of meiosis in *msh2-1* hybrids

To investigate the cause of reduced pollen viability and screen for possible meiotic phenotypes in the *msh2-1* hybrids, I generated a cytological atlas of meiosis. Meiotic chromosome spreads were performed on buds of floral stages 8 to 10 (Smyth et al., 1990), which are known to contain a range of meiotic stages (Armstrong & Jones, 2003), and the chromatin was stained with DAPI. The *msh2-1* mutation had no detectable phenotype when screening the meiotic stages of pachytene, diakinesis, metaphase I, dyad, and tetrad, for either the Col/Ler or Col/CLC hybrids (fig. 20A). The absence of a defect at pachytene indicates that loss of MSH2 function has no obvious effect on the homology search or synapsis in a hybrid genome.

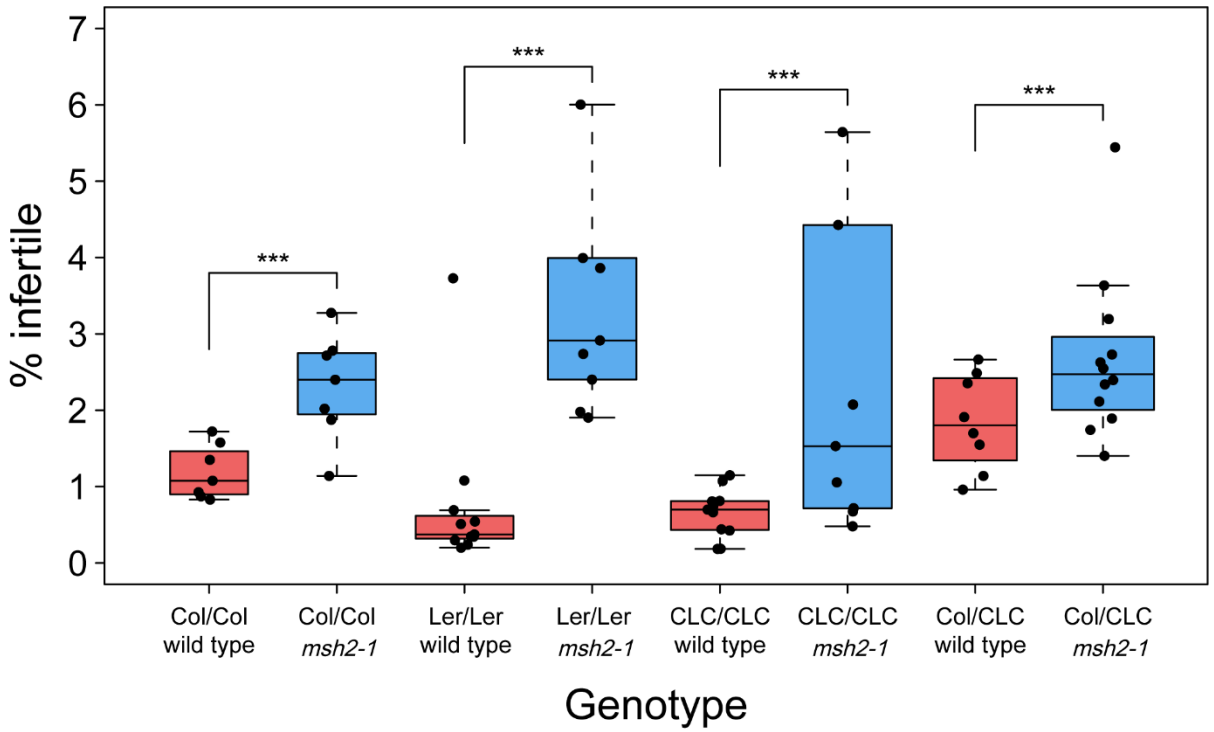
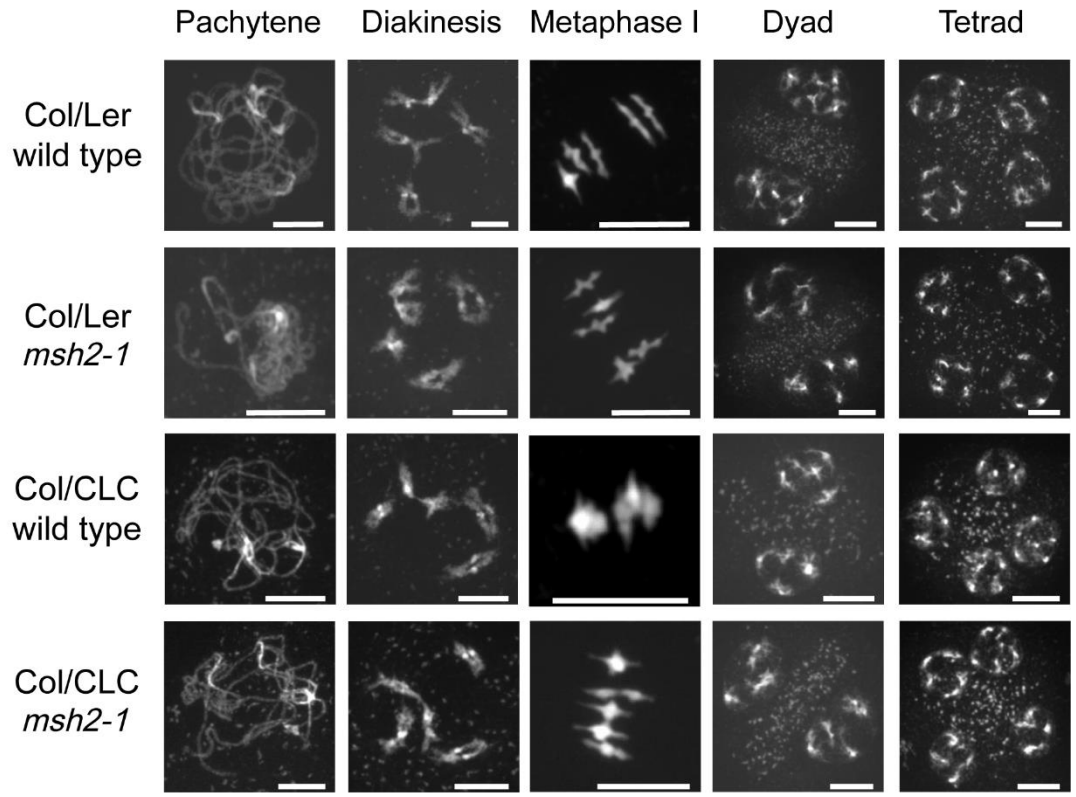


Figure 19. Pollen viability is decreased in the *msh2-1* mutant in inbred and hybrid contexts.

Boxplots of percentage pollen inviability measured by Alexander staining, in wild type (red) or *msh2-1* (blue) inbred (Col-0, L*er*-0 BC and CLC BC) and hybrid (Col/CLC) backgrounds. The box indicates the median and interquartile range, and the whiskers indicate the range (excluding outliers). Black dots represent replicate measurements. Statistical significance was tested using X^2 tests. *P* was < 0.0001 for all comparisons indicated.

However, I noted a striking difference in metaphase I morphology between the Col/Ler and Col/CLC hybrids. In comparison to Col/Ler, the metaphase I bivalents appeared more tightly associated in both the Col/CLC wild type and *msh2-1* spreads, and there was a significant reduction in the proportion of rod bivalents (largely representing single crossovers) and a corresponding increase in the proportion of ring bivalents (largely representing ≥ 2 crossovers) (*P* = 0.003 and 0.047, for wildtype and *msh2-1*, respectively; Wilcoxon rank sum test) (fig. 20B) (appendix 7.9 & 7.10) (López et al., 2012). These phenotypes are consistent with an increased crossover frequency in the Col/CLC hybrid (López et al., 2012), which is also consistent with both the FTL and CTL measurements in this study and the high crossover frequency previously reported in Col/CLC hybrids (fig. 16, 17, 18) (Ziolkowski et al., 2015). In contrast, I observed no significant difference in metaphase I morphology between the wild type and *msh2-1* genotypes of either Col/Ler or Col/CLC (*P* = 0.967 and 0.234, respectively; Wilcoxon rank sum test). This suggests that total crossover number likely remains unchanged, or is only subtly changed, in the *msh2-1* mutant.

A


Pachytene
Diakinesis
Metaphase I
Dyad
Tetrad
Col/Ler wild type
Col/Ler msh2-1
Col/CLC wild type
Col/CLC msh2-1

B

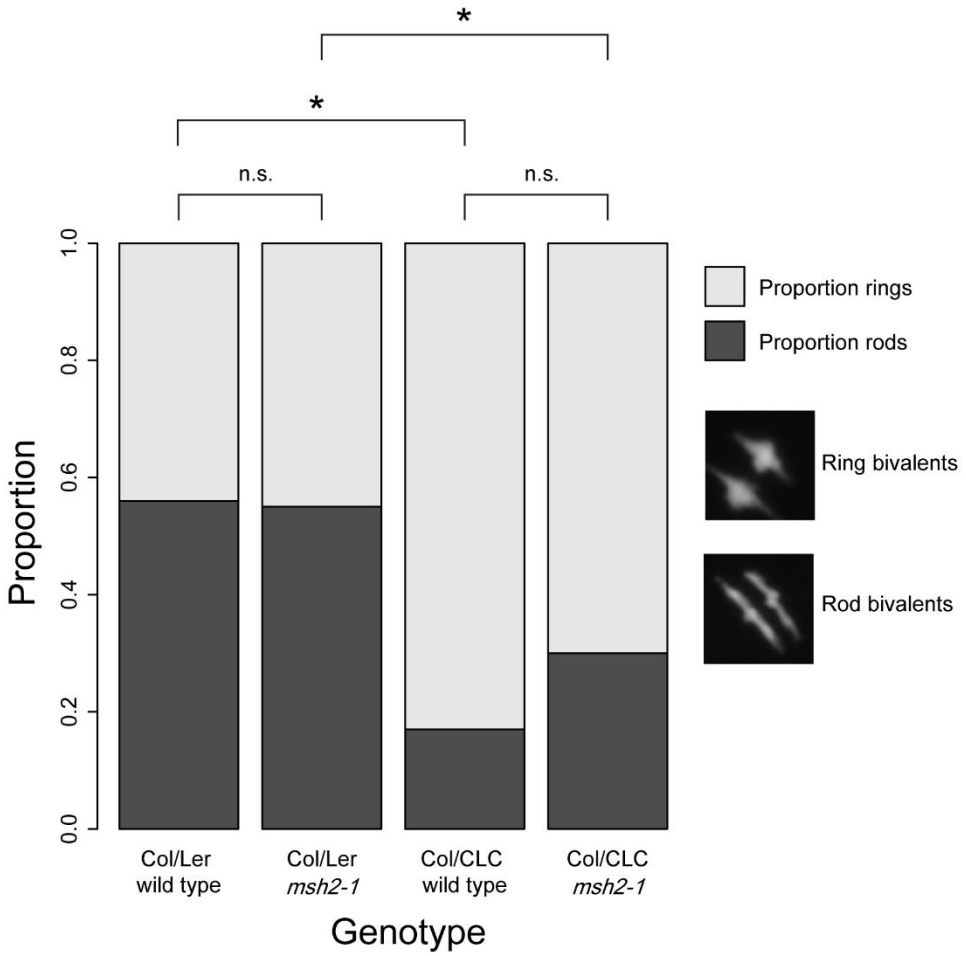

*
*
n.s.
n.s.
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
Col/Ler wild type
Col/Ler msh2-1
Col/CLC wild type
Col/CLC msh2-1
Genotype
Proportion rings
Proportion rods
Ring bivalents
Rod bivalents

Figure 20. The *msh2-1* mutation causes no meiotic defects at the cytological level.

(A) Representative DAPI spread images of pachytene, diakinesis, metaphase I, dyad and tetrad meiotic stages, for wild type and *msh2-1* Col/Ler and Col/CLC hybrids. Scale bar = 10 µm. (B) Proportion of rod (dark grey) and ring (light grey) metaphase I bivalents in wild type and *msh2-1* Col/Ler and Col/CLC hybrids. Two representative rod bivalents and two representative ring bivalents are shown. Statistical significance was tested using the Wilcoxon rank sum test, with the 'Benjamini & Hochberg' correction for multiple testing. Asterisks indicate $P < 0.05$; 'n.s.' indicates no significant difference.

3.8 Discussion

The changes in crossover frequency in the Arabidopsis *msh2-1* mutant were broadly consistent across the three genetic backgrounds analysed. Crossovers were significantly increased at the *I1b* genetic interval in two of the three hybrids tested, and were significantly increased at the *I2f* interval in both the hybrids tested. Across the two intervals, crossover frequency was increased by an average of 27.6% and 86.8% in the *msh2-1* Col/Ler and Col/CLC backgrounds compared to wild type, respectively. These increases are in general agreement with the previously observed *msh2-1* dependent increase in crossover frequency of 40% at a 2 Mb sub-telomeric interval on chromosome 5 in a Col/Ler hybrid (Emmanuel et al., 2006). However, based on these two intervals, it appears that crossover frequency in *msh2-1* counterintuitively increases by a greater proportion in the least divergent interval: *I1b* had an average divergence of 8.12 SNPs/kb, whilst *I2f* had a divergence of only 3.36 SNPs/kb. Furthermore, it appeared that diversity *per se* had no clear relationship with the crossover increases seen in *msh2-1* hybrids. Although the *I1b* interval in Col/Ct was slightly more divergent than in Col/CLC, there was a much smaller and non-significant crossover increase of 5.7%. However, such comparative conclusions are tentative due to the limited number of hybrid accessions and intervals analysed. Unlike the hybrids, crossover frequency at all tested genetic intervals remained unchanged in *msh2-1* in a Col-0 inbred context when compared to wild type.

In striking contrast to genetic intervals located in the chromosome arms, two overlapping intervals spanning the centromere showed significant decreases in crossover frequency in *msh2-1*. These ~7 Mb intervals contained the right pericentromere, centromere and approximately one half of the left arm of chromosome five. I initially predicted that these regions would show the greatest relative increase in crossover frequency in the *msh2-1* mutant hybrids, given the higher diversity of these regions (Borevitz et al., 2007; Ito et al., 2007; Alonso-Blanco et al., 2016). Although SNP density is only modestly increased for these intervals (e.g. 5.1 vs 3.1 SNPs/kb in *5.10* and *I2f*, respectively), this likely underestimates the diversity of these regions due to the difficulty of mapping short read sequences to these more repetitive genomic

regions, and hence identifying sequence variants. Furthermore, in addition to the 119 Mb Arabidopsis genome assembly, physical analyses indicate that a further 15 Mb of sequence that remain unassembled within the centromeric gaps (AGI, 2000; Hosouchi et al., 2002). The unassembled centromere regions are thought to consist largely of *CEN180* satellite repeats, which are also known to be diverse between accessions (Hosouchi et al., 2002; Borevitz et al., 2007; Ito et al., 2007). For example, sequence analysis of the tandemly repeated *CEN180* centromeric repeats indicated high levels of sequence divergence (ranging from 6–17%) between repeat monomers within the Col-0 genome (Simoens et al., 1988; Ito et al., 2007). Therefore, these observations are in opposition to those predicted from the literature: that MSH2 acts as a polymorphism dependent anti-crossover factor, with a greater anti-crossover activity on more divergent sequences.

Two further lines of evidence support my findings. Firstly, a pollen-typing assay at the *RESISTANCE TO ALBUGO CANDIDA1* (*RAC1*) crossover hotspot locus observed a significantly reduced crossover frequency in the Col/Ler *msh2-1* mutant relative to the wild type (Serra et al., 2018b). For this experiment, the L*er*-0 *msh2-1* BC line I generated was utilised by Dr Heïdi Serra. *RAC1* is located on the border of pericentromeric chromatin on chromosome 1, in a region of above average crossover frequency (Choi et al., 2016; Serra et al., 2018b). Furthermore, the *RAC1* locus contains a high level of Col-0 × L*er*-0 polymorphism: 27.4 SNPs/kb compared to a genome average of 3.85 SNPs/kb (Serra et al., 2018b). Hence, it was predicted that crossover frequency would increase in the *msh2-1* mutant at this locus, as the literature implicated polymorphism in repressing crossovers at endogenous hotspots (Cole et al., 2010; Drouaud et al., 2013; Serra et al., 2018b). Arabidopsis pollen-typing is analogous to sperm-typing methods, and allows the precise quantification of crossover frequency (Kauppi et al., 2009; Cole et al., 2010; Drouaud & Mézard, 2011; Choi et al., 2017). Briefly, genomic DNA was extracted from pollen of Col/Ler F_1 hybrids, which are heterozygous for polymorphisms across the *RAC1* locus (Drouaud & Mézard, 2011; Choi et al., 2017). Allele-specific PCR, using primers annealing to flanking SNP markers, then amplified a 9,482 bp region of the *RAC1* locus. These products are either single crossover (exchange of flanking markers) or parental molecules (non-exchange of flanking markers), depending on the primer combination. DNA samples are then diluted by titration, which allows precise determination of the concentrations of crossover and parental molecules. Surprisingly, when comparing the mean number of crossover and parental molecules per µl, genetic distance at *RAC1* was found to be significantly decreased from 0.093 cM in the wild type to 0.066 cM in the *msh2-1* mutant ($P = 2.94\times10^{-4}$; X^2 test) (fig. 21). As the *RAC1* locus also proved refractory to increases in additional hyper-recombination genetic backgrounds, such as the *recq4a recq4b* mutant and the *HEI10* overexpressor, this result may reflect a specific character of the locus, such as its

relatively high polymorphism density or centromere-proximal location (Serra et al., 2018b). However, the decrease at *RAC1* in the *msh2-1* mutant is also consistent with the reduced centromeric crossover frequency observed at the *5.10* and *5.11* intervals. Together with the results presented in this chapter, observations at *RAC1* argue for region-specific changes to crossover frequency in the *msh2-1* background.

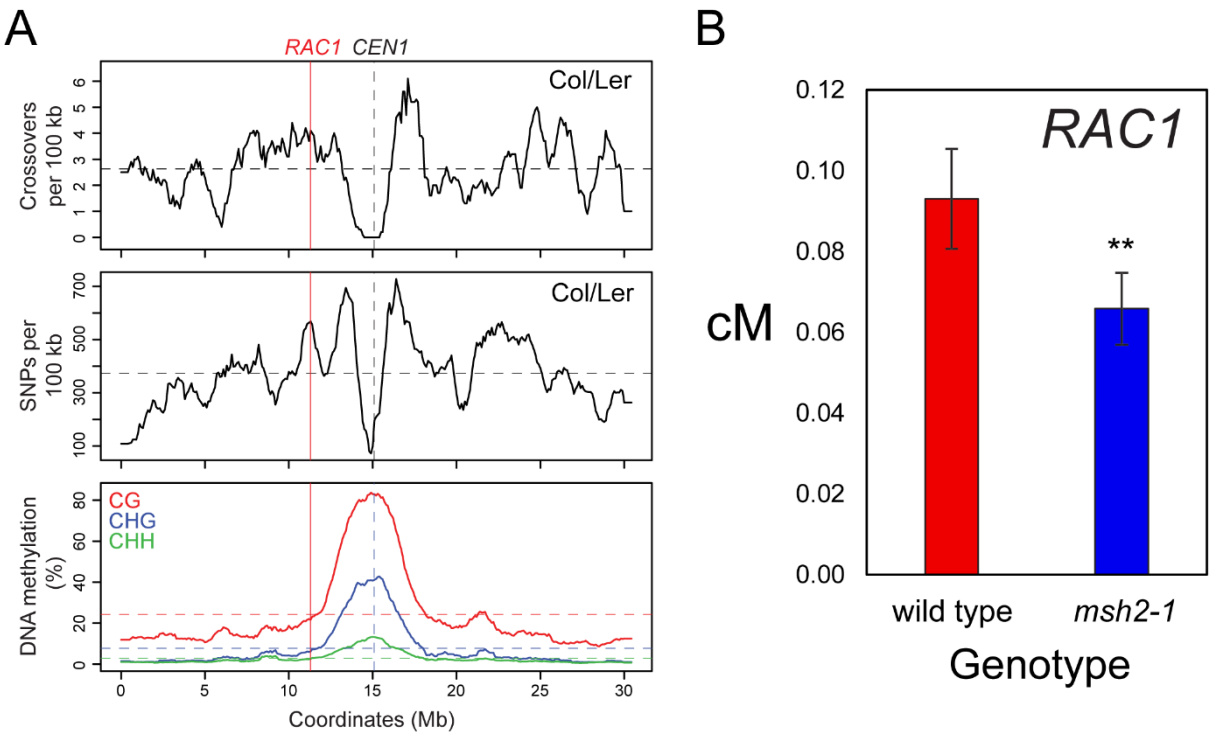


Figure 21. Crossover frequency in the *msh2-1* mutant is reduced at the *RAC1* crossover hotspot.

(A) Crossover frequency in a Col/Ler hybrid (crossovers/100 kb) (Serra et al., 2018a) (Upper panel), Col/Ler sequence divergence (SNPs/100 kb) (Zapata et al., 2016) (Middle panel), and % DNA methylation (CG red, CHG blue, CHH green) (Stroud et al., 2013) (Lower panel) are shown for Arabidopsis chromosome 1. Vertical dashed lines indicate the centromere. Mean values for the chromosome are indicated by horizontal dashed lines. The location of *RAC1* is indicated by a vertical red line. (Adapted from Serra et al., 2018b.) (B) Bar plots showing the genetic distance (cM) measured at the *RAC1* locus in wild type or *msh2-1* using pollen typing. Error bars represent standard deviation. A X^2 test was performed to assess statistical significance between the mean number of crossovers and parental molecules per µl ($P = 2.94\times10^{-4}$). (Adapted from Serra et al., 2018b.)

Secondly, an examination of regional differences in SNP and indel accumulation in wild type Arabidopsis lines indicated that mutations occur non-randomly across the genome (Ossowski et al., 2010). In a Col-0 wildtype, point mutation frequency is 44.1% higher in the pericentromeres versus non-pericentromeres (Ossowski et al., 2010). Deamination of methylated cytosines contributes to this trend, which is higher in the pericentromeres, in addition to ultraviolet light–induced mutagenesis (Ossowski et al., 2010). However, a study of mutation accumulation in an *msh2-1* MMR-deficient background found only a 5.9% increase of point mutations in the pericentromere relative to the chromosome arms, in comparison to wild type (Belfield et al., 2018). Interestingly, this study also found support for increased MMR activity within genes: point mutations within genes were increased by ~3-fold in the *msh2-1* mutation accumulation lines, relative to the wild type controls (Belfield et al., 2018). Together, these data suggest that MSH2-dependent MMR activity is least active over the pericentromeres, and greatest in the gene-rich chromosome arms.

The bias of MMR activity towards genes may be mediated by the Tudor domain of the Arabidopsis MSH6 protein, which binds strongly to the active chromatin mark H3K4me3 (Zhao et al., 2018). These observations in Arabidopsis are consistent with a study of mutation accumulation in human cancer cell lines, which also showed an MMR-dependent reduction in mutation rate within exons compared to introns (Frigola et al., 2017). These multiple lines of evidence suggest that loss of MMR function may in fact cause meiotic crossovers to redistribute across the genome due to the role of chromatin in shaping the localisation of MMR factors. Paradoxically, loss of MMR function may therefore cause crossovers to redistribute away from the more divergent centromeres, because loss of MMR-dependent crossover suppression is greater in the chromosome arms and hence causes a relatively greater increase crossover frequency in these regions.

Consistent with previous observations of decreased fertility in the *Solanum lycopersicum msh2* mutant (Sarma et al., 2018), pollen viability was decreased in the Arabidopsis *msh2-1* mutant, although to a much lesser extent. There was, on average, only a 1.86% increase in pollen inviability in *msh2-1*, which was also observed in all three inbred backgrounds (Col-0, L*er*-0 BC and CLC BC). Furthermore, I did not observe any defects in meiotic recombination in the *msh2-1* hybrids at the cytological level. Together, these results may support a 'haplosufficiency quality-checking' model, as previously proposed by Hoffman et al. (2004), where inviability in the *msh2-1* background is caused by exposed deleterious mutations in the haploid pollen generation. This model would explain the increased inviability in the inbred, as well as hybrid, backgrounds. This slight increase in pollen inviability is also consistent with the relatively low level of *de novo* mutations occurring in each *msh2-1* generation (Watson et al., 2016; Belfield et al., 2018). These are estimated at between ~30 (Watson et al., 2016) to ~100 *de novo*

mutations per somatic *msh2-1* generation (Belfield et al., 2018), based on deep sequencing and variant SNP calling after passaging the *msh2-1* mutation for several generations.

Although mutations accumulating in the *msh2-1* generation would lead to sequence polymorphisms being present between recombining chromosomes even for the Col-0 inbred FTL and CTL experiments, the sequence diversity would likely not be high given the low *de novo* mutation rate. Based on previous estimates (Watson et al., 2016; Belfield et al., 2018), I would predict the sequence divergence in the Col-0 experiments to be ~0.0007 SNPs/kb, approximately 10,000× lower than for the hybrid comparisons. Given this extremely low level of diversity, I believe it is still justified to refer to these lines as inbred. In support of this, the *fancm* mutation in a Nipponbare × Dongjin rice hybrid was found to increase recombination 2.3-fold relative to the wild type, despite being unable to increase crossover frequency in a hybrid context in Arabidopsis (Mieulet et al., 2018). However, the diversity between the recombining chromosomes was 0.09 SNPs/Kb, which the authors suggested may well be below a threshold sensitivity to polymorphism for meiotic recombination (Mieulet et al., 2018). However, I cannot discount the possibility that additional mutations have accumulated in the Col-0 *msh2-1* line during its passaging through haploid male gametophytic generations during backcrossing.

In addition to the unexpected regional changes in crossover frequency in the *msh2-1* background, the overall crossover frequency appeared elevated in the Col/CLC hybrid. Crossover frequency was higher in Col/CLC for all genetic intervals assayed, in comparison to Col/Ct or Col/Ler, and the proportions of bivalents with a ring morphology was higher, in comparison to Col/Ler. Together, these observations raise the question of how crossovers are distributed genome-wide in the Arabidopsis *msh2-1* mutant hybrids.

3.9 Acknowledgements

I would like to acknowledge Dr Christophe Lambing for providing advice on cytological procedures and analysis.

Chapter Four – Results – Genome-wide analysis of meiotic crossover in the *msh2-1* mutant

4.1 Introduction

Given the varying degree of change observed in crossover frequency across different genetic intervals in the *msh2-1* mutant hybrids, I next sought to investigate the effect of losing MMR function on crossovers genome-wide in two hybrid backgrounds (Col/Ler and Col/CLC). Genome-wide analyses of crossover frequency and distribution in *msh2* mutants has been performed previously in budding yeast (Martini et al., 2011; Oke et al., 2014; Cooper et al., 2018). For example, Class I crossovers are sensitive to polymorphism in budding yeast, as assessed using 1 or 2 kb windows, whilst crossovers in the *msh2* mutant are insensitive to polymorphism and a greater proportion form via the Class I pathway (Cooper et al., 2018). Consistent with analyses of yeast inter-species hybrids, loss of MMR increases crossover number in budding yeast hybrids (Hunter et al., 1996; Martini et al., 2011; Cooper et al., 2018). However, the budding yeast and Arabidopsis genomes show pronounced differences in several genomic parameters. For example, the Arabidopsis genome is 10× larger than the budding yeast genome (119.1 Mb versus 12.1 Mb), and has only 5 chromosomes compared to the 16 chromosomes in budding yeast (AGI, 2000; Duina et al., 2014). In Arabidopsis there is also significant epigenetic differentiation of euchromatin and heterochromatin, with megabase-scale regions of the genome, such as the centromere, showing high DNA methylation, condensed cytological structures and A/B compartment partitioning of the genome (Fransz et al., 2006; Stroud et al., 2013, 2014; Feng et al., 2014; Liu et al., 2016). In contrast, budding yeast has lost DNA methylation (Proffitt et al., 1984), and lacks the conserved heterochromatic histone modifications H3K9me2 and H3K9me3 (Bühler & Gasser, 2009). Furthermore, budding yeast centromeres are comprised of a single ~125 bp 'point' centromere, in contrast to megabase-scale heterochromatic and repetitive centromeres in Arabidopsis, that remain poorly assembled (Simoens et al., 1988; Hosouchi et al., 2002; Fransz et al., 2006; Borevitz et al., 2007; Ito et al., 2007). Transposable element (TE) densities are also higher in Arabidopsis, which comprise ~20% of the genome and are divided into multiple superfamilies of DNA and RNA elements (Slotkin & Martienssen, 2007; Underwood et al., 2017). In contrast, TEs comprise only 3.35% of the budding yeast genome and consist primarily of *Ty1-Ty5* LTR retroelements (Carr et al., 2012). In relation to MMR and meiotic recombination, Arabidopsis shows pronounced, megabase-scale regional variation in polymorphism density, in contrast to the relatively equally distributed polymorphism densities in budding yeast. Strikingly, whilst ~10 crossovers occur per meiosis in Arabidopsis hybrids,

~90 crossovers occur in budding yeast hybrids (Mancera et al, 2008; Wijnker et al., 2013), indicating a vastly increased cM per Mb in the budding yeast genome relative to Arabidopsis. Therefore, it is interesting to compare the *msh2* phenotype in these different genomic contexts.

In addition, whilst both Class I interfering and Class II non-interfering DNA repair pathways exist in Arabidopsis and budding yeast, the phenotypes of mutations in orthologous factors are not always identical. For example, the orthologous budding yeast *sgs1* and Arabidopsis *recq4a recq4b* mutations, both disrupting conserved DNA helicases, show different effects on total crossover numbers (Serra et al., 2018a). For example, deletion of *Sgs1* in budding yeast causes a relatively small 1.6-fold increase in crossover frequency (Jessop et al., 2006), and was also shown to function in a crossover resolution pathway (Zakharyevich et al., 2012). In contrast, the Arabidopsis *recq4a recq4b* mutations increase crossover frequency by 3.3-fold (Serra et al., 2018a), with additional crossovers forming via the Class II crossovers pathway (Séguéla-Arnaud et al., 2016; Serra et al., 2018a). Furthermore, the Class II crossover pathway in Arabidopsis appears more sensitive to interhomolog polymorphism (Girard et al., 2015; Fernandes et al., 2018a; Serra et al., 2018a), in comparison to budding yeast Class II crossovers (Cooper et al., 2018).

Interestingly, experiments in Arabidopsis showed a redistribution of crossovers into regions of polymorphism, at the megabase-scale, when heterozygous regions were juxtaposed with homozygous regions (Ziolkowski et al., 2015). Notably, this redistribution was dependent on the ZMM Class I crossover pathway (Ziolkowski et al., 2015). However, this megabase-scale observation, implying a positive relationship between heterozygosity and meiotic crossovers, appears inconsistent with effects of polymorphism observed at the fine-scale (Borts & Haber, 1987; Cole et al., 2010; Ziolkowski & Henderson, 2017; Serra et al., 2018b; Cooper et al., 2018). For example, local polymorphisms show a repressive effect on recombination, as measured at two endogenous crossover hotspots (Serra et al., 2018b), or using mitotic HR reporters (Li et al., 2004b; Opperman et al., 2004; Li et al., 2006). Taken together, these observations prove challenging to unify, but are likely caused by regulatory processes, including feedback mechanisms, acting at different scales and at different steps of the recombination reaction.

Therefore, to further assess the contribution of MMR to the relationship between crossover frequency and polymorphism, I utilised genotyping by sequencing (GBS) of recombinant F_2 plants, in order to generate high-resolution crossover maps in *msh2-1* mutant hybrids and their corresponding wild type hybrids.

4.2 Genotyping-by-sequencing of wild type and *msh2-1* F_2 recombinant populations

I performed GBS analysis on recombinant F_2 populations derived from hybrid F_1 parents, either with or without *MSH2* function (fig. 22). I selected hybrid Col/Ler and Col/CLC F_1 parents that were siblings of the *5.10* or *5.11* CTL hybrid populations (section 3.5), and genotyped for the *msh2-1* T-DNA and selected against the CTL fluorescent transgenes. The Col/Ler and Col/CLC hybrids were selected due to the significant changes in crossover frequency previously observed using FTL intervals in *msh2-1* mutants (section 3.4 & 3.5). Furthermore, Col/Ler hybrids have been previously used to generate genome-wide maps of crossovers in wild type and meiotic mutants (Giraut et al., 2011; Yelina et al., 2015; Fernandes et al., 2018a; Choi et al., 2018; Serra et al., 2018a; Underwood et al., 2018). Moreover, the genetically mosaic nature of CLC (section 3.2) – which includes a L*er*-0 chromosome 5 substitution, its varying polymorphism densities and inclusion of 'relict' sequences – made this an interesting genetic background to compare with Col/Ler hybrids in wild type and *msh2-1* (Alonso-Blanco et al., 2016).

Wild type and *msh2-1* mutant plants were self-fertilised, and genomic DNA was extracted from individual F_2 progeny and sequenced with an average of 1,747,241 reads, giving an average of 4.43× coverage per recombinant genome (appendix 7.11). The total reads from each genotype were aligned to the Arabidopsis reference genome using Bowtie2 (version 2.2.4; Langmead & Salzberg, 2012), in order to call sequence variants. Variant sites were then filtered to remove those with a quality score <100, >2.5× mean coverage, <10× reference coverage, and/or >100× variant coverage. Variant sites were repeat-masked within (a) TAIR10 transposable element annotations, (b) RepeatMasker output, (c) Tandem Repeats Finder output (Bensen et al., 1999), (d) Inverted Repeats Finder output (Warburton et al., 2004) and (e) crossover suppressed centromere regions (Copenhaver et al., 1999; Giraut et al., 2011), as previously described (Choi et al. 2013). The SNPs from these datasets were used to identify crossover sites in each F_2 individual. Briefly, sequence reads from each F_2 genome were independently aligned to the reference genome using Bowtie2, allowing mismatches. Aligned reads were compared to the dataset of filtered SNP variants generated from the total F_2 sequencing data, and the number of variant versus reference reads supporting each SNP site were obtained. These values were passed to TIGER (Trained Individual GenomE Reconstruction), a computational pipeline for reconstructing the mosaic genomes of recombinant individuals from low-coverage sequencing data (Rowan et al., 2015; Yelina et al., 2015). Briefly, the genotype along each chromosome was estimated by counting read alignments at the defined SNP variant positions that supported either the Col-0 (A) or variant (L*er*-0 or CLC (B)) parental alleles, using a sliding window of 10,000 markers (Rowan et al., 2015). TIGER output files provide the coordinates of transitions between A/AB/B genotypes,

which are assigned to the midpoint of the 5′ and 3′ SNPs that flank the point of genotype change. These are termed ‘crossover midpoints’ and were used for all subsequent GBS crossover analyses, unless otherwise stated. Experimental support for this approach to mapping crossovers has been provided via direct confirmation of crossover positions using Sanger sequencing (Rowan et al., 2015), by comparison with crossover maps generated using an alternative link-read sequencing approach (Sun et al., 2018), and from expectations based on previous genetic maps (Copenhaver et al., 1998; Salomé et al., 2012; Choi et al., 2013; Wijnker et al., 2013; Yelina et al., 2015).

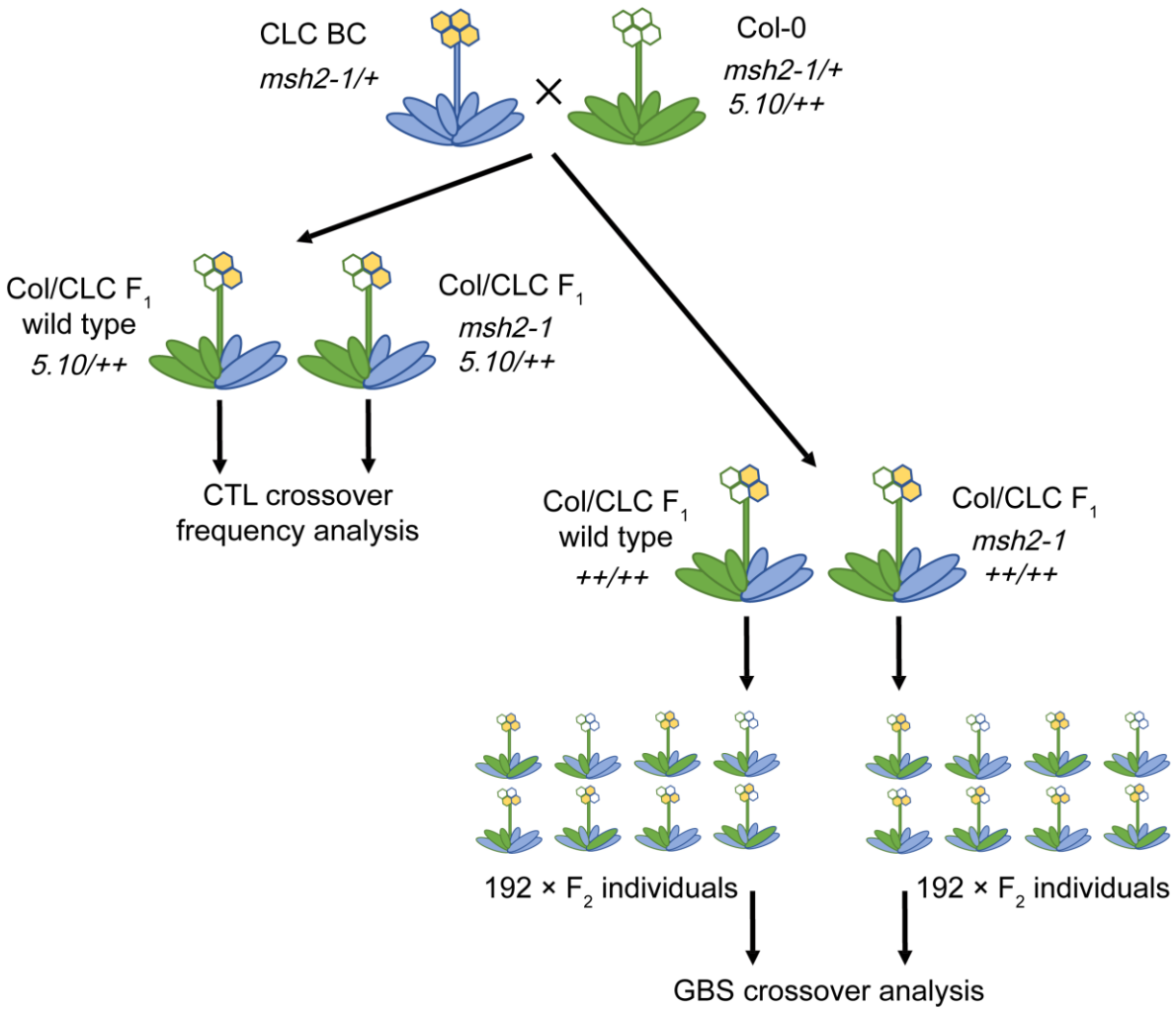


Figure 22. Diagram of crossing strategy used to generate GBS F_2 recombinant populations.

Col-0 *msh2-1*/+ *5.10*/++ plants (green and white) were crossed to either L*er*-0 BC *msh2-1*/+ or CLC BC *msh2-1*/+ plants (blue and yellow). The CLC background is chosen as representative. The resulting F_1 hybrids (half green and white, half blue and yellow), either with or without the *5.10* reporter transgenes, were genotyped as either wild type or *msh2-1* mutants. These plants were then self-fertilised to produce recombinant F_2 progeny. Those F_2 populations containing the *5.10* reporter were scored for crossover frequency at this interval, and 192 wild type and 192 *msh2-1* F_2 individuals from the non-*5.10* F_1 parents were processed for GBS sequencing library construction.

This analysis resulted in the reconstruction of 192 individual F_2 genotypes, for Col/Ler *msh2-1*, Col/CLC wild type and Col/CLC *msh2-1*. However, filtering for low genome coverage (<0.3×), a single duplicated individual present as a result of errors in library preparation, and noisy allele-frequency traces that are indicative of poor read mapping and/or contamination during library preparation reduced these sets to 187, 189, and 191, for Col/Ler *msh2-1*, Col/CLC wild type and Col/CLC *msh2-1*, respectively (fig. 23). As a Col/Ler wild type control, I utilised a previously published dataset of 240 Col/Ler wild type F_2 individuals (Serra et al., 2018a).

Analysis of the SNP variants defined using the Col/Ler and Col/CLC *msh2-1* mutant total F_2 data revealed a 4.55 and 4.23 Mb region of Col-0 homozygosity surrounding the *msh2-1* T-DNA introgression on chromosome 3, respectively. The boundaries of these regions were identified based on recurrent crossover coordinates at these positions, identified in independent F_2 libraries, caused by the presence of Col-0-to-L*er*-0 genotype transitions, which also corresponded to an absence of SNPs identified through variant calling using the total F_2 sequencing data. Crossovers called by TIGER due to the junction between Col-0 and Ler-0 or CLC sequence either side of these regions were masked from the dataset. These masked regions spanned 3.0–9.0 Mb and 4.0–9.5 Mb of chromosome 3 for Col/Ler and Col/CLC, respectively. An identical region was masked from the corresponding wild type datasets. An additional introgression of Col-0 sequence on the right arm of the CLC chromosome 2 was similarly masked from the Col/CLC wild type and *msh2-1* mutant datasets, from 6.5–10.3 Mb.

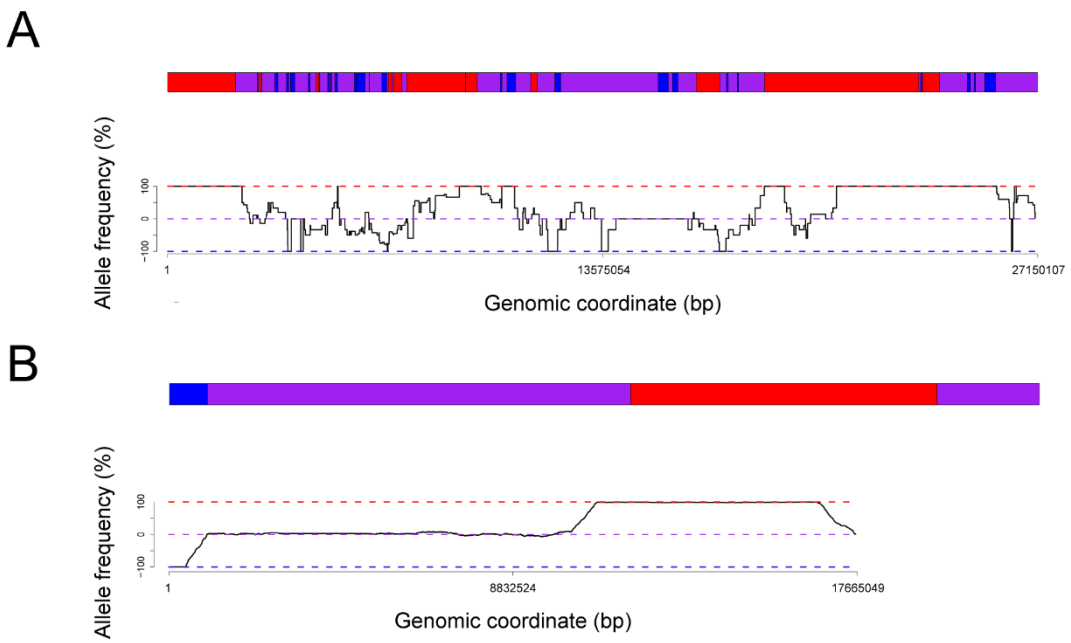


Figure 23. Examples of low and high quality GBS allele frequency traces.

(A) An example of a low quality genotype reconstruction using the TIGER pipeline, due to low sequence coverage. The upper panel represents the points of genotype exchange, between homozygous Col-0 (red), heterozygous (purple), and homozygous L*er*-0 (blue) regions. The lower panel shows the allele frequency (in %) relative to the Col-0 reference genotype, for a portion of the chromosome. (B) The same as (A), but shown for a high quality genotype reconstruction.

Several additional Col-0 regions were identified from the GBS sequencing data, which were present in a subset of the F_1 parents due to incomplete introgression of the *msh2-1* T-DNA from the Col-0 background. These were identified due to the repeated identification of crossovers with identical midpoints in F_2 plants that shared parents. For Col/Ler, the first 170 kbs of chromosome 4 and a 200 kb region between 7.8–8.0 Mb of chromosome 4 retained Col-0 sequences. Therefore, crossovers at these junctions were masked from the wild type and *msh2-1* datasets. For Col/CLC, five additional regions were identified as containing residual Col-0 introgressions. Firstly, one region on chromosome 1 were identified, between 3.46–4.65 Mb. Secondly, two regions on chromosome 2 were identified between 4.47–5.39 Mb and 16.12–18.88 Mb. Finally, one region on chromosome 4 was identified, between 3.13–4.80 Mb. In all cases, the regions spanning these junctions were masked from the wild type and *msh2-1* mutant datasets. The final datasets contained 1,718, 1,394, 1,495 and 1,534

crossovers in the Col/Ler wild type, Col/Ler *msh2-1*, Col/CLC wild type and Col/CLC *msh2-1*, respectively (appendix 7.12). Together, these datasets assigned crossover positions to the midpoints of SNP intervals, with a mean SNP interval width of 1,031, 1,268, 728, and 1,038 base pairs, respectively.

4.3 Total crossover number remains unchanged in the Col/Ler and Col/CLC *msh2-1* mutants

First, I assessed the total number of crossovers per F_2 in the wild type Col/Ler and Col/CLC populations. I observed that mean crossover number was 10.5% higher in the Col/CLC wild type compared to the Col/Ler wild type (7.91 vs 7.16 COs/F_2, respectively; $P = 3.33\times10^{-4}$; Wilcoxon rank sum test). This was consistent with the chiasma analysis of metaphase I chromosome spreads, where the proportion of ring:rod bivalents increased from 0.45 in the Col/Ler hybrids to 0.77 in the Col/CLC hybrids (fig. 20B), and with previously published observations of high crossover frequency in the CLC background (Ziolkowski et al., 2015). However, this GBS analysis likely under-represents the total crossover number because the Col-0 region of chromosome 3 surrounding the *msh2-1* T-DNA is masked in both the Col/Ler and Col/CLC populations, which hence masks ~5% of the genome from crossover detection. Moreover, a further 5.8% of the genome is masked in the Col/CLC population due to the presence of two Col-0 introgressed regions.

Secondly, I compared the total crossover number per F_2 between the wild type and *msh2-1* populations, for Col/Ler and Col/CLC. This revealed small increases in total crossover number, from 7.16 to 7.29 and 7.91 to 8.03 per F_2, in the two hybrid contexts, respectively (fig. 24). Crossovers increased by 4.1% in the Col/Ler and 1.5% in the Col/CLC *msh2-1* mutants, relative to their corresponding wild types. However, neither of these increases were statistically significant ($P = 0.0534$ and 0.602, respectively; Wilcoxon rank sum test). These increases are far less than those seen in studies of crossover frequency in budding yeast wild type and *msh2* hybrids, such as the S288C×SK1 (~0.57% divergence) and the S96×YJM789 hybrids (~0.6% divergence), where crossover frequency was observed to increase by 40% and 25% in *msh2* mutant hybrids compared to the wild type hybrids, respectively (Oke et al., 2014; Cooper et al., 2018). However, although the level of sequence divergence in the current study is lower than those used in budding yeast studies, with 0.48% divergence for Col/Ler and 0.55% divergence for Col/CLC, it is surprising that this difference would cause a different outcome in terms of crossover phenotype (Martini et al., 2011; Cooper et al., 2018).

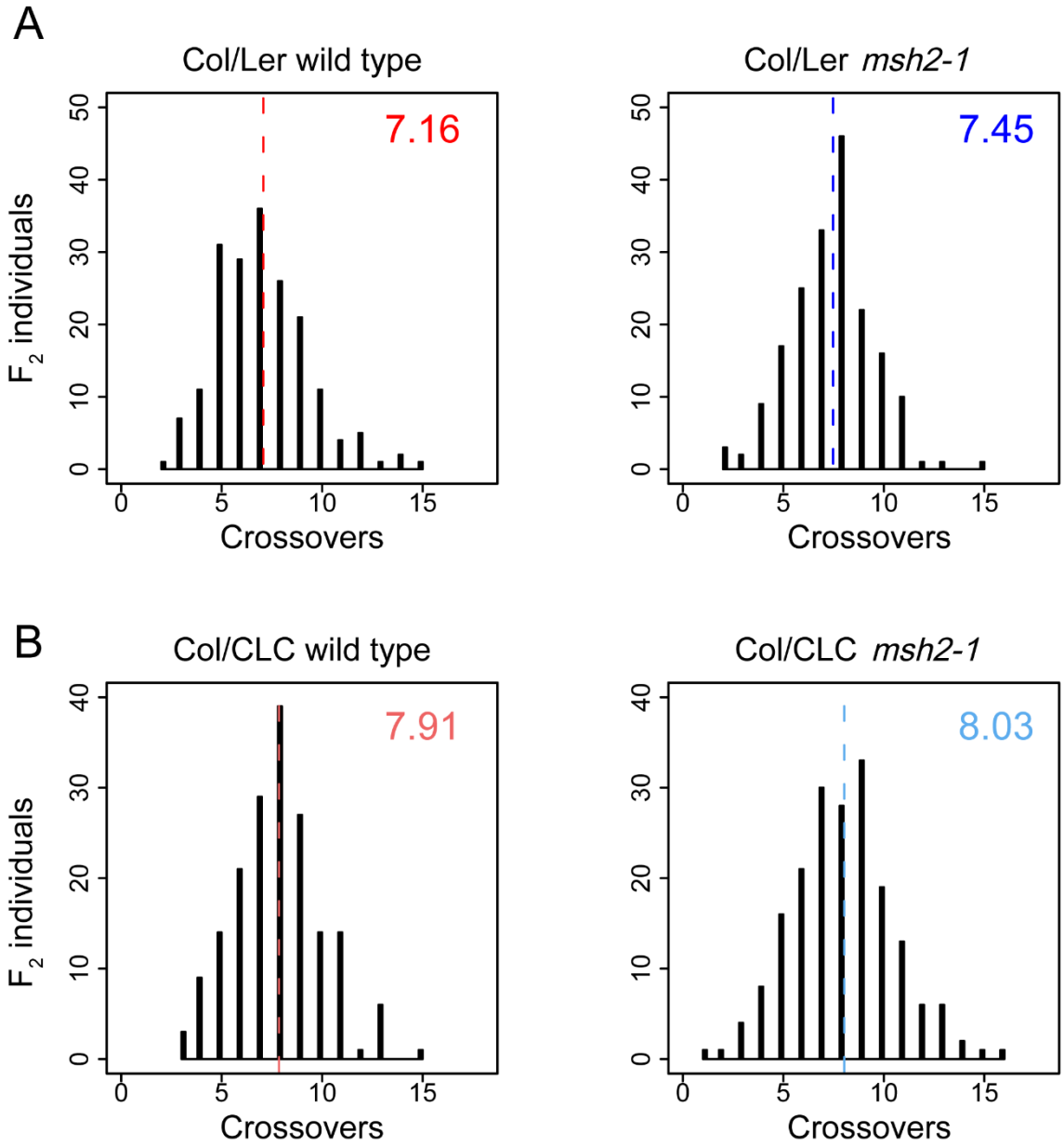


Figure 24. Total crossover frequency remains unchanged in the *msh2-1* mutant.

(A) Histograms indicating the frequency of Col/Ler F_2 individuals with a given number of crossovers, for wild type (left panel) and *msh2-1* mutant (right panel) populations. The mean number of crossovers per F_2 individual is indicated with a vertical dashed line and is stated in the top right corner, for wild type (red) and *msh2-1* (blue). Col/Ler wild type data are from Serra et al. (2018a). (B) As described in (A), but for the Col/CLC hybrid. Statistical comparisons were performed with the Wilcoxon rank sum test. No significant difference was found between the wild type and *msh2-1* genotypes.

4.4 A *trans*-modifier of meiotic recombination alters the crossover landscape of CLC's L*er*-0 chromosome five

I sought to compare the crossover distributions between Col/Ler and Col/CLC wild types. To plot crossover frequency along the chromosomes, crossover coordinates were pooled for each genotype and analysed using a 300 kb sliding window. A comparison of the wild type crossover landscapes between Col/Ler and Col/CLC revealed striking differences (fig. 25A). Elevated crossover frequency was observed on the boundary of the pericentromere in both F_2 populations, although Col/CLC displayed a greater elevation in crossover frequency over these regions. 26.4% of crossovers occurred in the pericentromeres of the Col/Ler wild type, whilst a significantly greater amount (35.0%) occurred over the pericentromeres in the Col/CLC wild type ($P = 1.45\times10^{-4}$; X^2 test) (table 3). However, for both hybrids the centromeres remained relatively crossover supressed, in contrast to the adjacent pericentromeres. Notably, despite the Col/CLC hybrid containing a substituted L*er*-0 chromosome 5 (fig 12B), this chromosome displayed a crossover landscape more closely resembling the other four Col/CLC chromosomes than the Col/Ler chromosome 5 (fig. 25A). For example, 26.9% of crossovers occurred in the pericentromere of Col/CLC chromosome 5, which was greater than the 23.9% occurring over the pericentromere of Col/Ler chromosome 5. However, this increase was not statistically significant ($P = 0.244$; X^2 test), likely due to a lack of power caused by the lower crossover numbers observed per chromosome compared to genome-wide. However, this may provide evidence for *trans* acting modifiers from the other CLC chromosomes modifying the crossover landscape on chromosome 5, particularly in the pericentromeric regions, as the *cis* pattern of polymorphism is the same as in Col/Ler.

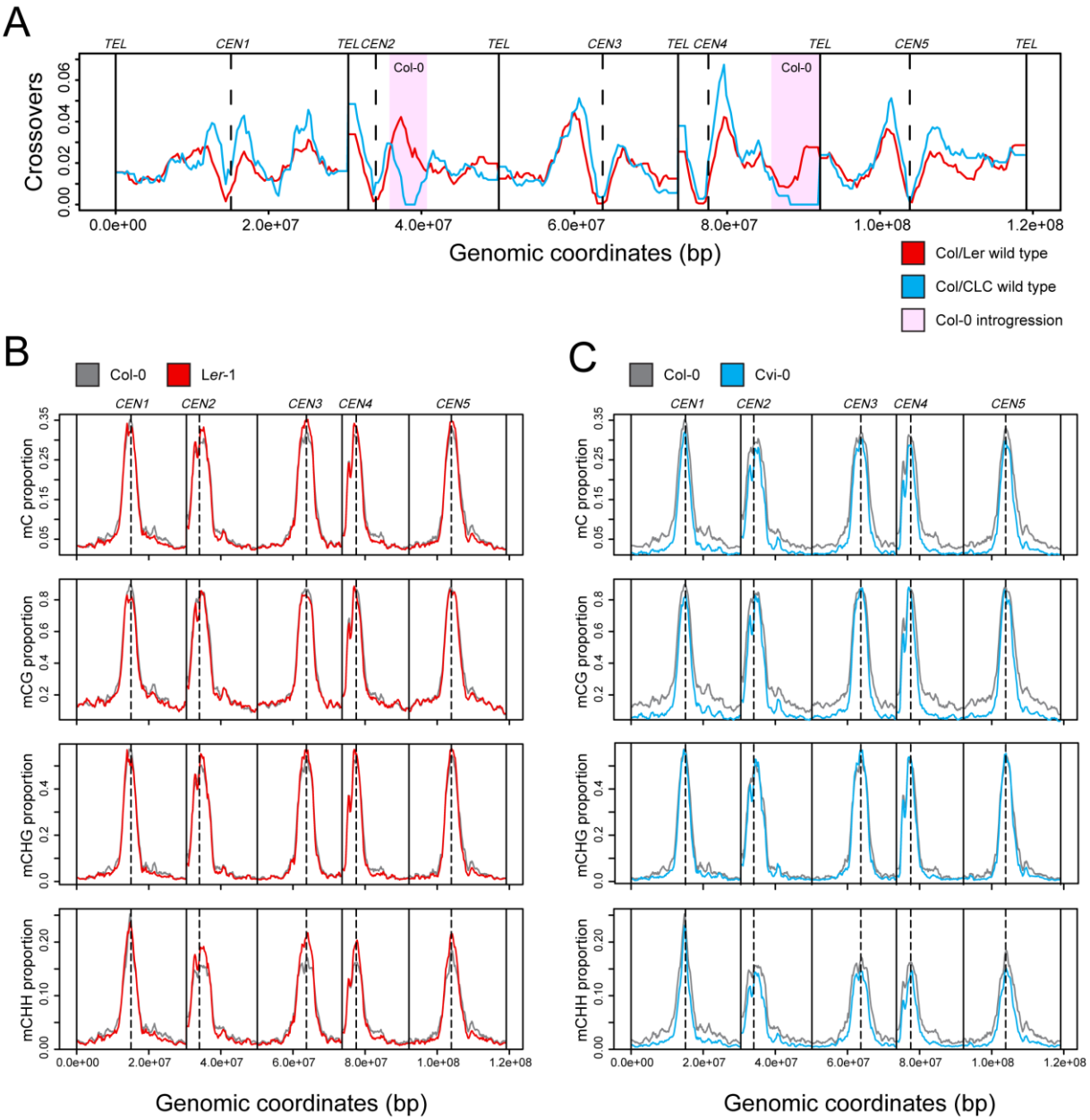


Figure 25. Genome-wide crossover distributions in Col/Ler and Col/CLC wild type hybrids and analysis of DNA methylation landscapes.

(A) Crossover frequency is plotted for Col/Ler (red) and Col/CLC (pale blue) wild types. Crossover frequency was analysed in 300 kb windows and normalised by the number of F_2 individuals in each population. All five chromosomes are represented on a continuous x-axis, from 0 to 119,146,348 bp. The positions of telomeres (*TEL*) are indicated by vertical lines and of centromeres (*CEN1–CEN5*) with vertical dashed lines. The location of two Col-0 regions in the CLC background are indicated with pink boxes. Col/Ler wild type data are from Serra et al. (2018a). (B) The proportion of cytosine methylation in Col-0 (grey) or L*er*-1 (red) wild types at each position in the genome was plotted in 10 kb windows, for all contexts, CG, CHG and CHH contexts. Methylation data are from Kawakatsu et al. (2016). (C) As for (B) but for Col-0 (grey) and Cvi-0 (pale blue) wild types.

Col/Ler wild type (*n* = 240)													
Region	**Chr1**	**CO/F$_2$**	**Chr2**	**CO/F$_2$**	**Chr3**	**CO/F$_2$**	**Chr4**	**CO/F$_2$**	**Chr5**	**CO/F$_2$**	**Total COs**	**Total COs/F$_2$**	***P*-value**
Arms	335	1.40	213	0.89	173	0.72	228	0.95	315	1.31	**1,264**	**5.27**	**-**
Pericentromeres	99	0.41	107	0.45	79	0.33	70	0.29	99	0.41	**454**	**1.89**	**-**
Centromeres	0	0	0	0	0	0	0	0	0	0	**0**	**0**	**-**
Total	**434**	**1.81**	**320**	**1.33**	**252**	**1.05**	**298**	**1.24**	**414**	**1.73**	**1,718**	**7.16**	
Col/Ler *msh2* (*n* = 187)													
Arms	303	1.62	200	1.07	159	0.85	184	0.98	288	1.54	**1,134**	**5.94**	**0.069**
Pericentromeres	59	0.32	70	0.37	32	0.17	30	0.16	69	0.37	**260**	**1.36**	**5.47×10^{-5}**
Centromeres	0	0	0	0	0	0	0	0	0	0	**0**	**0**	**-**
Total	362	1.94	267	1.44	191	1.02	214	1.14	357	1.91	**1,394**	**7.29**	
Col/CLC wild type (*n* = 189)													
Arms	279	1.48	136	0.72	110	0.58	154	0.81	283	1.50	**962**	**5.09**	**0.017**
Pericentromeres	122	0.65	95	0.50	101	0.53	100	0.53	105	0.56	**523**	**2.77**	**1.45×10^{-4}**
Centromeres	4	0.02	0	0	1	0.01	2	0.01	3	0.02	**10**	**0.05**	**0.002**
Total	**405**	**2.14**	**231**	**1.22**	**212**	**1.12**	**256**	**1.35**	**391**	**2.07**	**1,495**	**7.91**	
Col/CLC *msh2* (*n* = 191)													
Arms	339	1.77	148	0.77	187	0.98	165	0.86	311	1.63	**1,150**	**6.02**	**0.008**
Pericentromeres	89	0.47	62	0.32	57	0.30	78	0.41	95	0.50	**381**	**1.99**	**8.67×10^{-6}**
Centromeres	0	0	0	0	1	0.01	1	0.01	1	0.01	**3**	**0.02**	**0.09**
Total	**428**	**2.24**	**210**	**1.10**	**245**	**1.28**	**244**	**1.28**	**407**	**2.13**	**1,534**	**8.03**	

Table 3. Crossovers counts in centromeres, pericentromeres and chromosome arms for wild type or *msh2-1* genotypes, in Col/Ler and Col/CLC F$_2$ populations.

Crossovers were assigned to either the chromosome arms, pericentromeres or centromeres, for each genotype. Centromeres were defined as regions surrounding the centromeric assembly gaps that are suppressed for meiotic crossover, pericentromeres were defined as regions adjacent to the centromere with higher than average DNA cytosine methylation (Yelina et al., 2015; Underwood et al., 2018). The arms were defined as the remaining chromosome sequences. Col/Ler wild type data are from Serra et al. (2018a). Statistical significance between wild type and *msh2-1* genotypes, or between the two hybrid wild types, was assessed with X^2 tests.

Notably, a total of 10 crossovers occurred within the centromeres of the Col/CLC wild type, whilst none occurred in the centromeres of the Col/Ler wild type, which corresponded to a statistically significant increase (P = 0.002: X^2 test) (table 3). However, one caveat for this analysis is that crossover coordinates are being assigned relative to the Col-0 reference genome assembly. Hence, these centromeric crossovers should be treated with caution, due to our poor understanding of centromere structure in both Col-0 and CLC backgrounds, of which three centromeres are derived from the Cvi-0 genome. This is particularly relevant given our knowledge of centromere structural diversity in Cvi-0, relative to other Arabidopsis accessions, which includes clusters of variant centromeric satellites (Ito et al., 2007). In relation

to this, it is also important to consider the likely reduced accuracy of TIGER crossover assignment in proximity to the centromere, due to the difficult of mapping within repetitive regions and hence the sparsity of markers. However, it should be noted that crossovers were identified in the centromeres of all five chromosomes in the Col/CLC wild type, including 3 crossovers in the centromere of the chromosome 5 L*er*-0 introgression (table 3). This suggests that these crossovers represent real events, because in contrast no crossovers were observed across the centromere of chromosome 5 in the Col/Ler wild type or *msh2-1* crossover maps. However, as centromere coordinates are defined using the Col-0 genome assembly, it remains possible that crossovers assigned to the remaining centromeres (i.e. centromeres 1-4) in the Col/CLC maps are an artefact due to our lack of knowledge of structural variation within the Cvi-0 centromeres. Furthermore, it is also possible that substantial epigenomic reprogramming occurs in the Col/CLC hybrids, which may alter the epigenetically-defined component of centromere identity and thereby make regions of the Col-0 centromere crossover competent.

As I observed a difference in pericentromeric crossovers between the Col/Ler and Col/CLC wild type maps, I hypothesized that euchromatin/heterochromatin structure may differ between these lines. Therefore, I collaborated with Ms Pallas Kuo to quantify the proportion of heterochromatin in Col/Ler and Col/CLC wild types and to investigate whether a change in heterochromatin content might correlate with changes in pericentromeric and centromeric recombination. Specifically, we quantified the area of heterochromatin of DAPI-stained pachytene spreads, as defined based on pixel intensity values. The DAPI-dense nucleolus organiser regions form loops at the ends of chromosomes, and were excluded from the quantification analysis. We then calculated the percentage of each meiocyte's total area occupied by heterochromatin (appendix 7.13). This analysis revealed that heterochromatin occupied 9.0% of the area in Col/CLC, which was significantly greater than the 6.2% in Col/Ler (fig. 26) ($P = 1.083\times10^{-5}$; Wilcoxon rank sum test). Whilst this observation is consistent with either greater total levels of heterochromatin or more diffuse heterochromatin, I believe it likely indicates the latter. For example, this is consistent with total DNA methylation levels being lower in the Cvi-0 accession relative to Col-0 (Kawakatsu et al., 2016; Picard & Gehring, 2017), and the observation that heterochromatic chromocenters are less pronounced and numerous in Cvi-0 mitotic cells, relative to L*er*-0 (Tessadori et al., 2009).

To directly address this question, the proportion of methylated cytosines in Col-0, L*er*-1 and Cvi-0 was plotted along the chromosomes in 10 kb sliding windows, using previously published bisulphite sequencing data (fig. 25B,C) (Kawakatsu et al., 2016). Methylated cytosines were assessed in the symmetric CG content, and the non-symmetric CHG and CHH contexts (where H indicates any base other than G). As the L*er*-0 accession was absent from the bisulphite dataset, the L*er*-1 accession was used for comparison, which is a direct descendent from the

L*er*-0 stock. One caveat to this analysis is that methylation data were generated from the inbred accessions, and may not be representative of the hybrid context. This was a particular concern for Col/CLC, which has a mosaic genetic background primarily derived from Cvi-0. However, whilst the L*er*-1 methylation was highly similar to Col-0, for all sequence contexts, Cvi-0 showed pronounced reductions in CG and CHH contexts (fig. 25B,C). Total methylation levels were lower in Cvi-0, which was most noticeable in the chromosome arms (fig. 25C). Interestingly, methyl-cytosine reductions in Cvi-0 were most pronounced for the CG and CHH contexts, whilst CHG context methylation remained similar to Col-0 levels (fig. 25C). Together with the cytological quantification, this analysis suggests that altered DNA methylation levels may be contributing to the differences in meiotic crossover landscape between Col/Ler and Col/CLC, consistent with roles for DNA methylation in shaping the meiotic recombination landscape in Arabidopsis (Yelina et al., 2012; Yelina et al., 2015; Choi et al., 2018; Underwood et al., 2018).

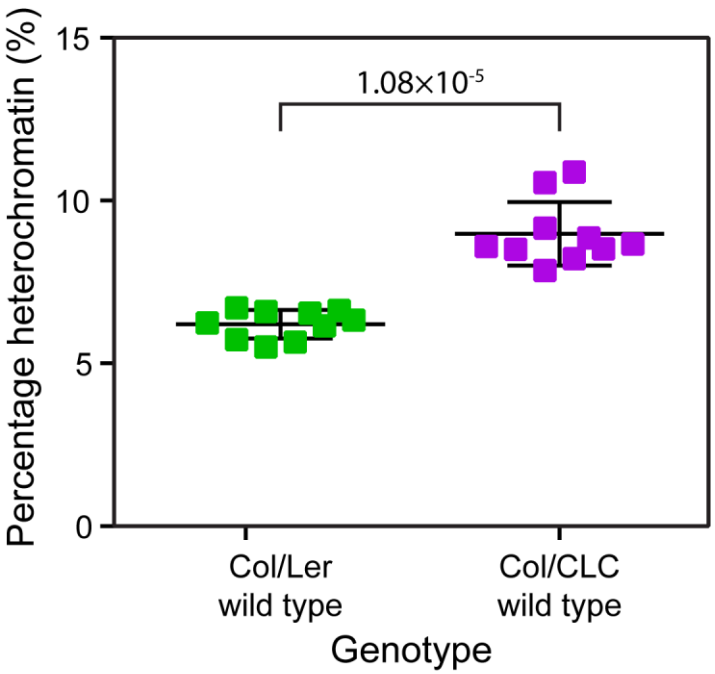


Figure 26. Compared to Col/Ler hybrids, heterochromatin area is increased in Col/CLC hybrids.

Heterochromatin area as a percentage of each meiocytes total area was defined based on pixel intensity values for each image of DAPI-stained pachytene spreads, for Col/Ler wild type (green squares) and Col/CLC wild type (purple squares) genotypes. Each square represents the percentage of heterochromatin in a single meiocyte. Mean and interquartile range are indicated with horizontal black lines. Statistical significance was assessed using the Wilcoxon rank sum test, and the *P* value is indicated.

4.5 The *msh2-1* mutation causes crossovers to redistribute away from the diverse pericentromeres

We previously observed crossover frequency increases in the chromosome arms and decreases over the pericentromere in *msh2-1* hybrids, using FTL, CTL and pollen typing genetic intervals (section 3.4 & 3.5) (Serra et al., 2018b). Therefore, I sought to compare crossover patterns throughout the genome for wild type and *msh2-1* populations, by plotting pooled crossover coordinates in 300 kb sliding windows. This revealed a subtle pattern of crossover redistribution in the *msh2-1* mutant, where crossovers were decreased in the pericentromeres and increased within the chromosome arms (fig. 27). I quantified crossover numbers in either the chromosome arms, pericentromeres or centromeres, for each of the four genotypes (table 3). Compared to the wild type, this revealed a 29.4% and 29.0% reduction in crossovers in the pericentromeres relative to the other genomic regions, in the Col/Ler and Col/CLC *msh2-1* mutant, respectively ($P = 5.47\times10^{-5}$ and 8.67×10^{-6}, respectively; X^2 test). In contrast, I observed a corresponding 10.6% and 16.5% increase in the proportion of crossovers occurring in chromosome arms in the Col/Ler and Col/CLC *msh2-1* mutants, relative to wild type. However, this increase was only statistically significant for the Col/CLC comparison ($P = 0.069$ and 0.008, respectively; X^2 test).

To visualise genome-wide trends, I plotted crossover frequency along the telomere to centromere axis, averaged across all chromosome arms after scaling chromosomes for differences in their length and normalising for the number of F_2 individuals in each population (fig. 28A). This confirmed the trend observed at the chromosome scale, where crossovers were reduced close to the centromere and slightly increased along the chromosome arms in *msh2-1*. To address the statistical significance of differences in crossover frequency between wild type and *msh2-1*, Fisher's exact tests were performed on 2×2 contingency tables of crossover counts inside and outside a series of scaled chromosome windows. Counts for each scaled window were summed across all chromosome arms. *P*-values and multiple-testing-corrected *P*-values were plotted along the telomere to centromere axis (fig. 28B). This confirmed that crossovers were significantly reduced close to the centromeres, and significantly increased in the interstitial, central region of the chromosome arms, for both the Col/Ler and Col/CLC *msh2-1* mutants.

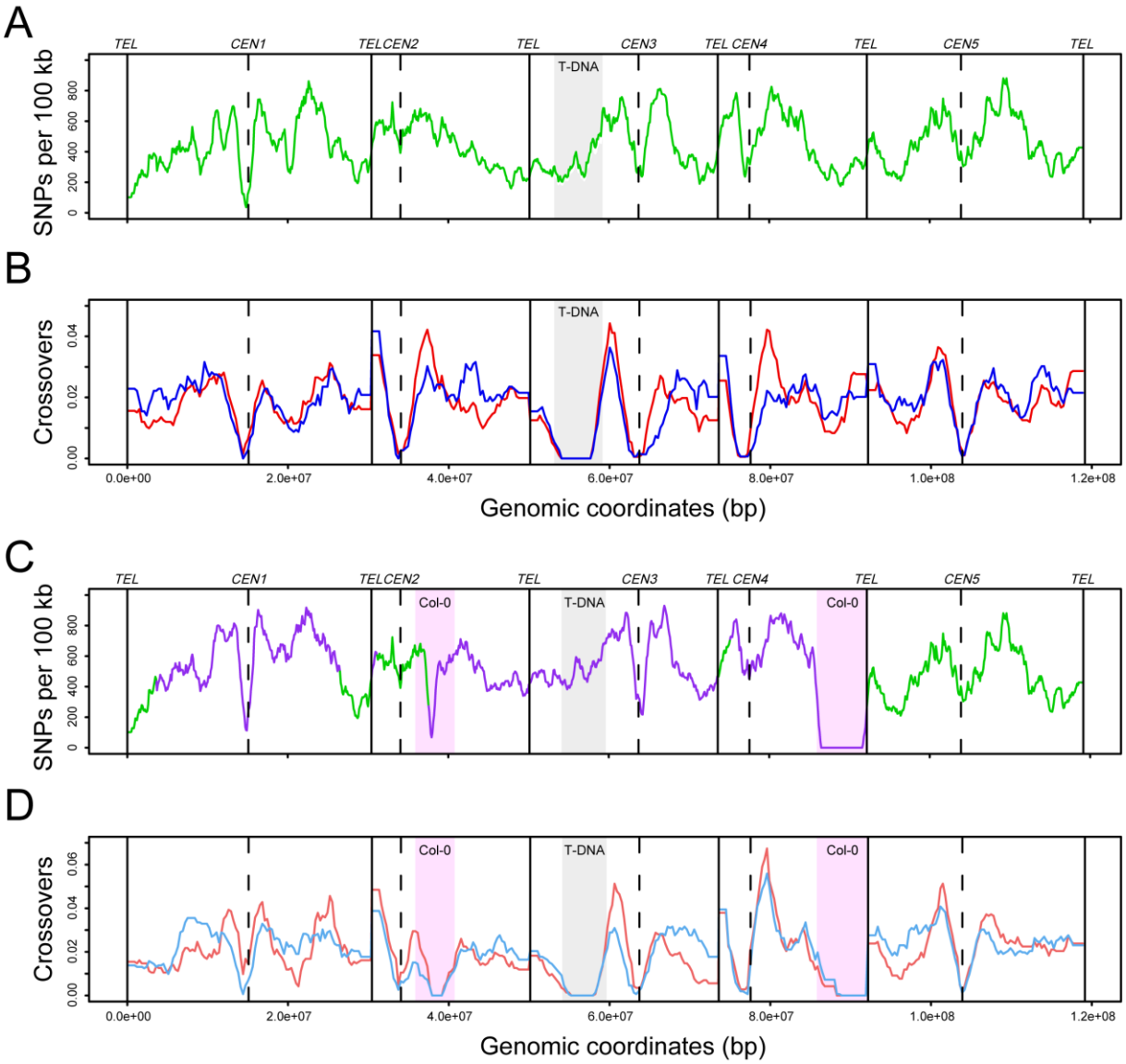


Figure 27. Genome-wide crossovers distributions in the Col/Ler and Col/CLC *msh2-1* mutants.

(A) Col/Ler SNP density (green line) is plotted along the genome, in 100 kb sliding windows. The grey box indicates the region of Col-0 sequence surrounding the introgressed *msh2-1* T-DNA. The positions of telomeres (*TEL*) are indicated by vertical lines and of centromeres (*CEN1–CEN5*) with vertical dashed lines. (B) As for (A) but with Col/Ler wild type (red) or *msh2-1* mutant (blue) crossover frequency, plotted in 300 kb sliding windows. Crossovers are normalised by the total F_2 number in each population. (C) As for (A) but with Col/CLC SNP density. Cvi-0 SNPs are indicated in purple and L*er*-0 SNPs are indicated in green. The location of two Col-0 regions in the CLC background are indicated with pink boxes. (D) As for (B) but with Col/CLC wild type (pale red) and *msh2-1* mutant (pale blue) crossovers. Col/Ler wild type data are from Serra et al. (2018a). SNP densities were calculated from the 1,135 genomes datasets (Alonso-Blanco et al., 2016).

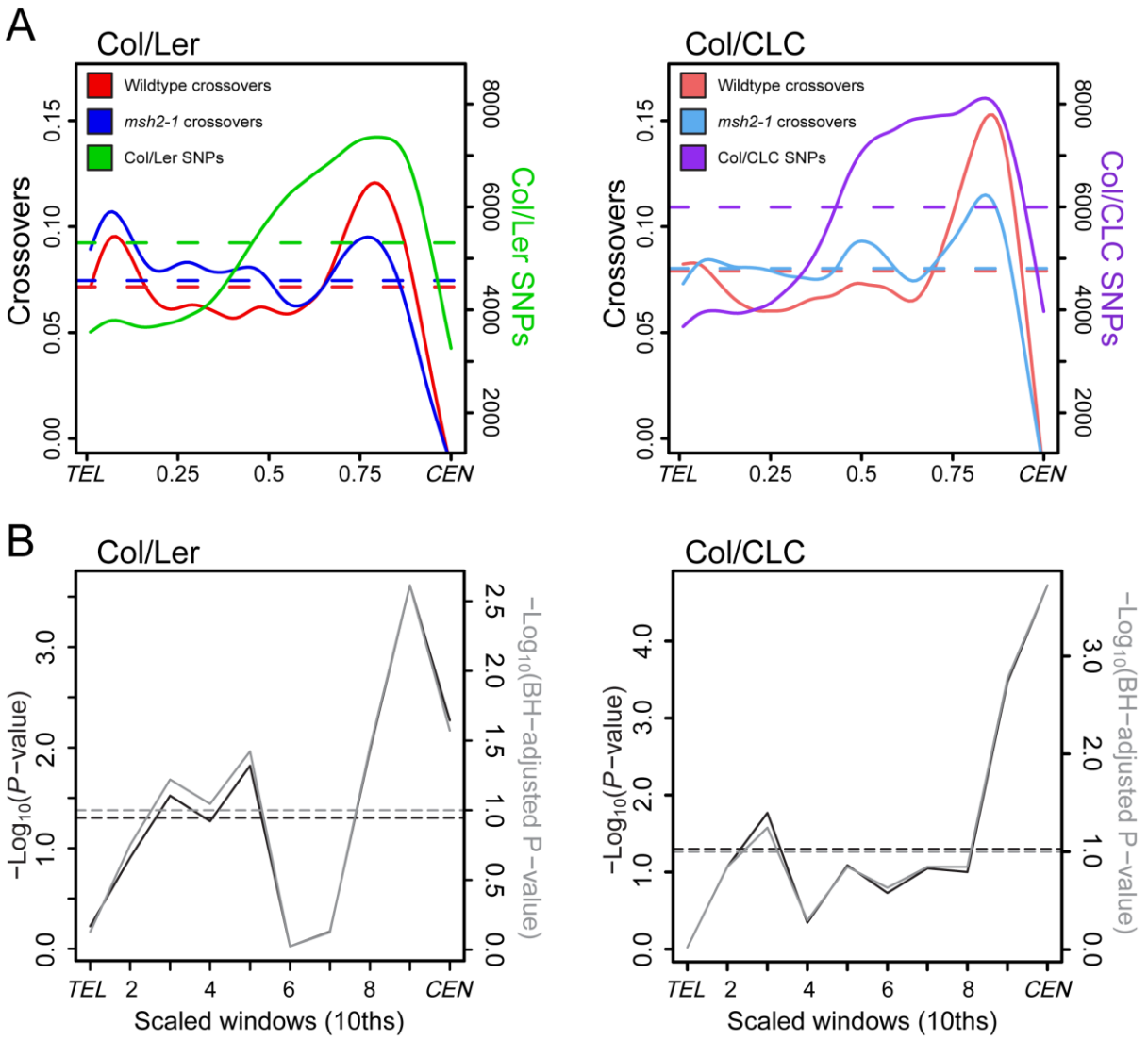


Figure 28. Crossovers redistribute along the telomere–centromere axis in the *msh2-1* mutant.

(A) Col/Ler wild type (red) and *msh2-1* mutant (blue) crossovers and Col/CLC wild type (pale red) and *msh2-1* mutant (pale blue) crossovers are plotted along a normalised telomere (*TEL*) to centromere (*CEN*) chromosome axis. The corresponding Col/Ler (green) and Col/CLC (purple) SNP datasets are shown (Alonso-Blanco et al., 2016). Crossovers are normalised by the total F_2 number in each population. Col/Ler wild type data are from Serra et al. (2018a). (B) A statistical analysis was performed to assess the statistical significance between Col/Ler wild type and *msh2-1* mutant (left) and Col/CLC wild type and *msh2-1* mutant (right) crossover distributions. The crossover profiles along the normalised *TEL–CEN* axis were divided into 10 adjacent windows of equal size, and *P*-values and multiple-testing-corrected *P*-values were derived from Fisher's exact tests performed on 2×2 contingency tables of crossover counts inside and outside of each scaled window. *P*-values were adjusted using the 'Benjamini–Hochberg' (BH) method. P-values (black) and adjusted-*P*-values (grey) were -$\log_{10}$ transformed and plotted along the *TEL–CEN* axis. Horizontal dashed lines correspond to an unadjusted *P*-value threshold of -$\log_{10}(0.05)$ (black) and adjusted-*P*-value threshold of -$\log_{10}(0.1)$ (grey). Values above these thresholds correspond to intervals with statistically significant differences.

Surprisingly, these changes in crossover distribution in the *msh2-1* mutant indicated that crossovers were decreasing in the regions with greatest interhomolog polymorphism, which is the opposite of what is expected if MSH2 is acting as a mismatch-dependent anti-crossover factor (fig. 28A), as observed in budding yeast (Cooper et al., 2018). This was clearly visualised by plotting SNP density along the telomere to centromere axis, in a manner analogous to that described above. This revealed that the point of highest SNP density at the centromere-proximal sequences showed a significant reduction in crossover frequency in *msh2-1*, at windows 0.8-1.0 and 0.9-1.0 of the proportional length in Col/Ler and Col/CLC, respectively ($-\log_{10}$(adjusted *P*) > 1.0; Fisher's exact test). In contrast, crossovers were significantly increased in the chromosome arms of Col/Ler and Col/CLC *msh2-1* mutants, in windows 0.3-0.6 and 0.3 of the proportional chromosome length, respectively ($-\log_{10}$(adjusted *P*) > 1.0; Fisher's exact test). This relationship was most pronounced in the Col/Ler hybrid, where SNP density dropped more rapidly outside of the pericentromeres (fig. 28A,B). In the Col/CLC hybrid, SNP density remained higher outside of the pericentromeres, and crossovers were only significantly increased further towards the ends of the chromosomes (window 0.3) (fig. 28B). Strikingly, these relationships suggest that crossovers in a MMR-deficient background are redistributing away from the divergent regions of the genome, counter to our initial expectations.

In order to validate the changes in the GBS crossover landscapes, I sought to compare these results with the previous FTL and CTL analyses. I calculated crossover frequency in centimorgans for the *I1b*, *I2f* and *5.10* intervals from the GBS crossover data. For the Col/Ler comparisons, the fluorescent interval measurements and the GBS crossovers were in close agreement for the *I1b* (increasing by 25.9% and 39.5%, respectively) and *5.10* intervals (decreasing by 9.2% and 4.7%, respectively) (fig. 29A). However, in contrast to the FTL data, the GBS crossover data indicated a reduction in crossover frequency at *I2f* in *msh2-1*, relative to the wild type (fig. 29A). For the Col/CLC hybrid, the two methods were again in close agreement at the *5.10* interval (decreasing by 4.2% and 1.0%, respectively) (fig. 29B). In contrast, although overall crossover frequency calculated from the GBS data was reduced at both the *I1b* and *I2f* intervals, relative to the FTL data, both intervals showed consistent increases in the *msh2-1* mutant. This comparison indicates that crossover changes in the *msh2-1* mutant are broadly consistent between the different methodologies, although there are significant discrepancies.

There are two possible explanations for the discrepant intervals. Firstly, both *I1b* and *I2f* are of a small size (1.85 and 0.67 Mb, respectively), relative to the size of the Arabidopsis genome (fig. 12A). This may make them inappropriate for comparison due to the small number of GBS crossovers occurring within these intervals. For example, 17 and 10 crossovers were observed within *I2f* in the Col/Ler wild type and *msh2-1* populations, respectively. Therefore, there is a

lack of sufficient power to use the GBS crossovers to make comparisons at such narrow physical intervals. In contrast, thousands of crossover events are observed per genotype within these genotypes with the fluorescent reporter system. Secondly, it is noteworthy that for both hybrids the *5.10* CTL interval was consistent with the GBS data. Not only is this interval larger (6.7 Mb), but the seed-based CTL data are an average of the crossover frequency during male and female meiosis. In Arabidopsis, male and female genetic maps vary greatly in size and distribution (Giraut et al., 2011; Fernandes et al., 2018a). For example, genetic maps are 575 cM and 332 cM in Arabidopsis male and female meiosis, respectively (Giraut et al., 2011; Fernandes et al., 2018a). Furthermore, male crossovers are elevated towards the chromosome ends, which is the location of *I1b* and *I2f*, whereas female crossover frequency is low (Giraut et al., 2011; Fernandes et al., 2018a). This is likely to further contribute to the differences observed between the FTL and GBS data at the *I1b* and *I2f* intervals. In addition, this phenomenon may in part explain the more modest increase in crossover frequency in the Col/CLC wild type GBS data, relative to the Col/Ler. Previous comparisons of crossover frequency measurements in Col/Ler and Col/CLC backgrounds have been predominantly made using pollen-based (i.e. male meiosis) FTL intervals, and were extremely high for Col/CLC (Ziolkowski et al., 2015). In comparison, the sex-averaged *420* interval, located on the left arm if chromosome 3, in fact showed a decrease in the Col/CLC wild type hybrid (Ziolkowski et al., 2015). These data may suggest a high degree of sexual dimorphism in meiotic recombination rate in the Col/CLC background.

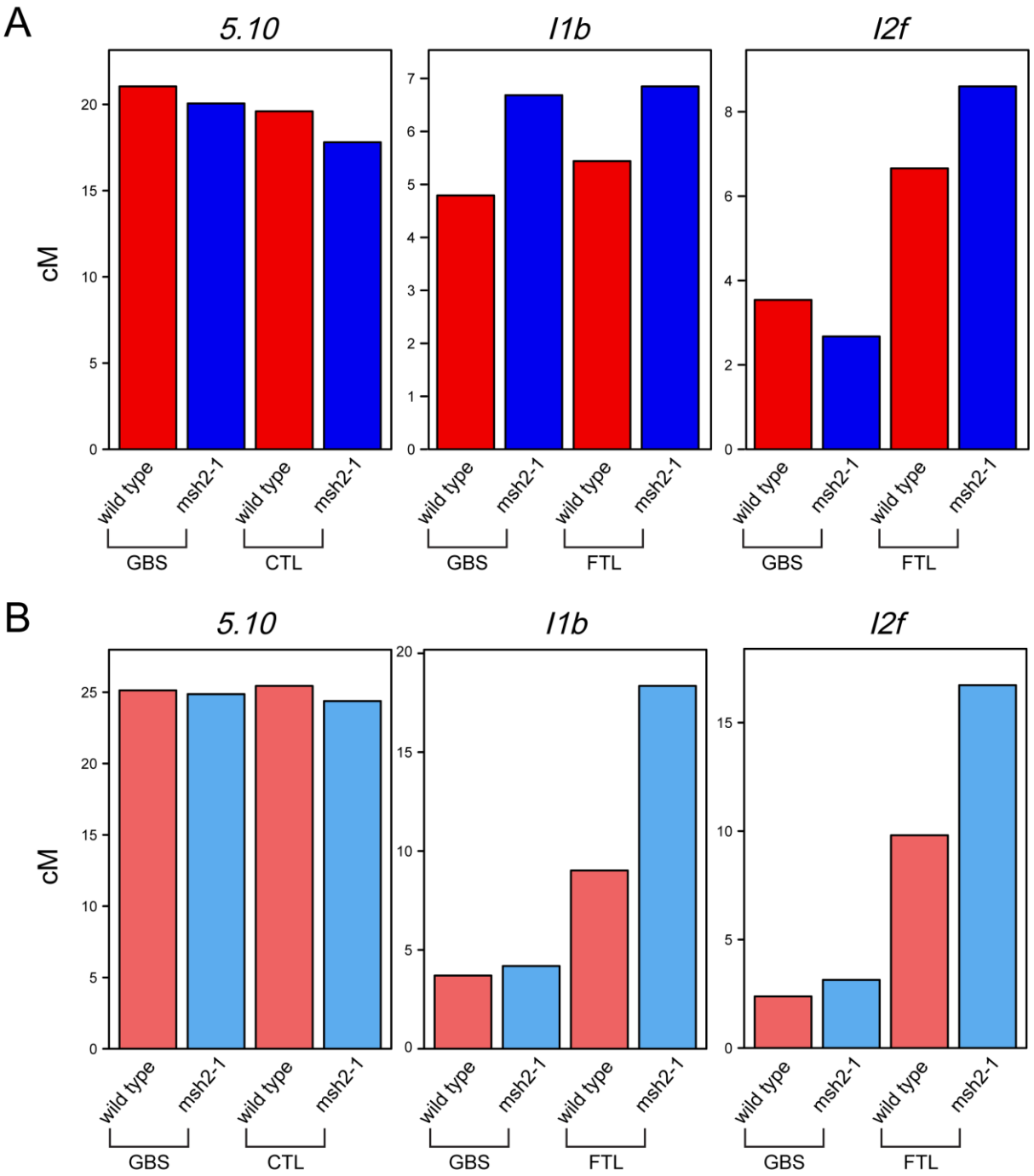


Figure 29. A comparison of crossover frequency at FTL and CTL intervals, calculated from fluorescent crossover measurements or GBS crossover data.

(A) Crossover frequency in centimorgans (cM) at the CTL interval *5.10* and the FTL intervals *I1b* and *I2f*. Crossover frequency in the Col/Ler wild type (red) or *msh2-1* mutant (blue) was calculated using the single-interval data or the GBS data. (B) As described in (A) but with the Col/CLC wild type (pale red) and *msh2-1* mutant (pale blue) data.

4.6 Polymorphism density is reduced around crossovers in *msh2-1* compared to wild type, across multiple scales.

The redistribution of crossovers from the pericentromeres to the chromosome arms in the *msh2-1* mutant hybrids indicated that recombination is reduced in areas of higher polymorphism, at the chromosome scale. Therefore, I sought to investigate this relationship at multiple physical scales. At the chromosome scale, correlations were made between crossovers and SNPs densities, calculated using adjacent 300 kb windows. For this analysis, regions of the L*er*-0 or CLC BC genomes containing Col-0 introgressions – such as the Col-0 T-DNA regions or the two Col-0 regions in CLC – were excluded from both the crossover and SNP datasets. Both Col/Ler and Col/CLC wild type crossovers were weakly positively correlated with SNP density (Pearson's r = 0.36 and 0.47, respectively; $P = 3.58\times10^{-12}$ and $<2.2\times10^{-16}$, respectively), consistent with previous observations of historical and experimental crossover frequency in Arabidopsis (Salomé et al., 2012; Choi et al., 2013; Ziolkowski et al., 2015; Serra et al., 2018a). I also assessed the chromosome scale relationship between crossovers and polymorphism using additional independent unpublished crossover maps generated from sequencing F_2 populations from Col-0 × Bur-0 and Col-0 × Ws-4 wild type F_1 hybrids, which were kindly provided by Dr Christophe Lambing and Dr Emma Lawrence. These hybrids are on average 0.47% and 0.35% diverged, respectively. Crossovers were positively correlated with interhomolog polymorphism for the Col/Bur hybrid (Pearson's r = 0.21; $P = 1.91\times10^{-5}$) which again confirmed this positive association within Arabidopsis (fig. 30A). Surprisingly, crossovers in the Col/Ws hybrid showed no correlation with polymorphism, in contrast to the other three hybrids (Pearson's r = -0.03; P = 0.068) (fig. 30B). The lack of positive correlation for the Col/Ws hybrid may be due to the reduced statistical power in these datasets, as both Col/Bur and Col/Ws maps are derived from only 96 recombinant F_2 individuals, in contrast to the 240 and 189 individuals in the Col/Ler and Col/CLC wild type datasets, respectively. However, the level of interhomolog polymorphism for the Col/Ws was the lowest of all the hybrids analysed, perhaps suggesting that a positive genome-wide association between crossovers and diversity requires a minimum degree of sequence divergence.

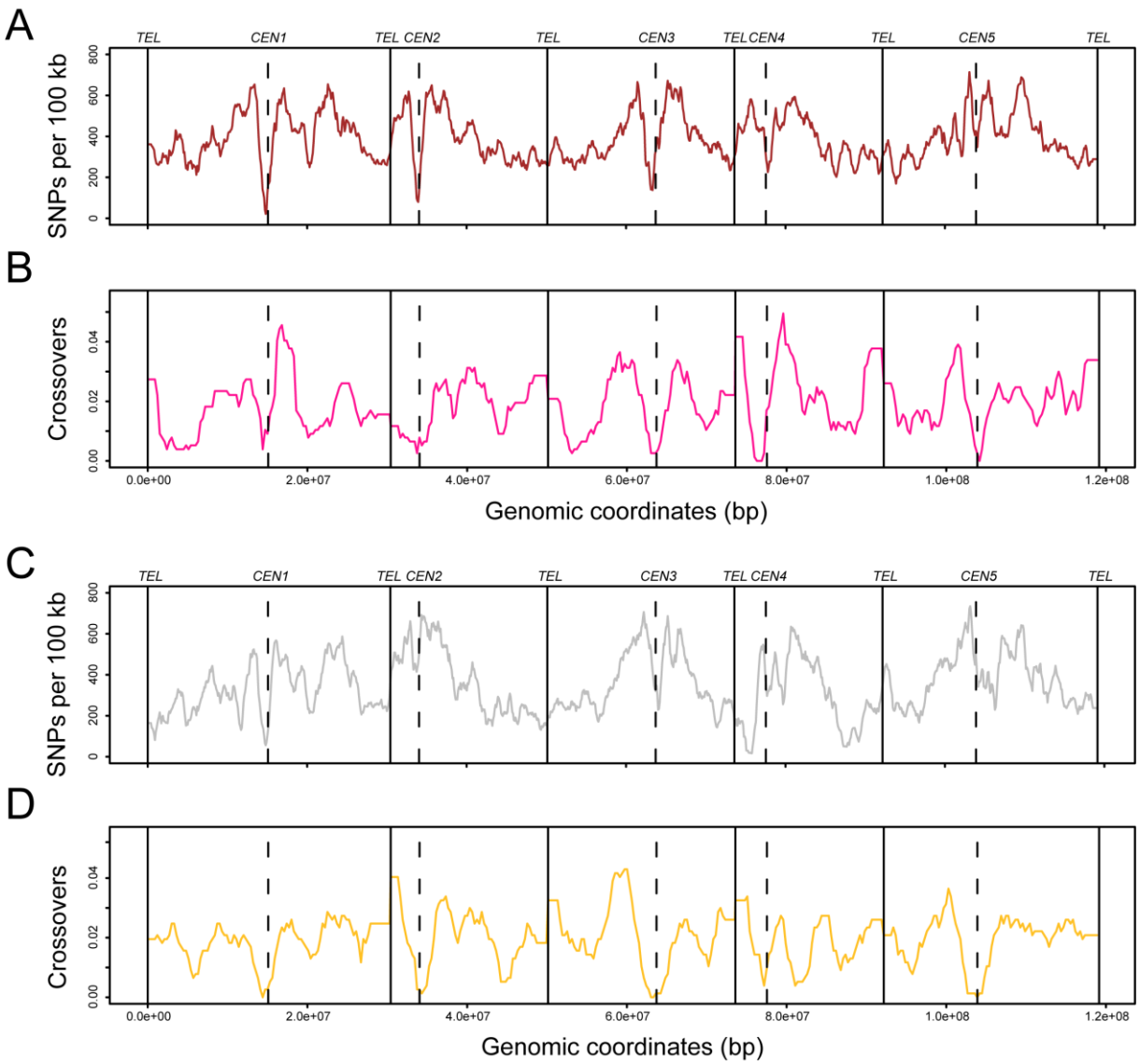


Figure 30. Genome-wide crossovers maps in the Col/Bur and Col/Ws wild type hybrids.

(A) Col/Bur SNP density (brown line) is plotted along the genome, in 100 kb sliding windows. The positions of telomeres (*TEL*) are indicated by vertical lines and of centromeres (*CEN1–CEN5*) with vertical dashed lines. (B) As for (A) but with Col/Bur wild type crossover frequency (pink line), plotted in 300 kb sliding windows. Crossovers are normalised by the total F_2 number in each population. (C) As for (A) but with Col/Ws SNP density (grey line). (D) As for (B) but with Col/Ws wild type crossover frequency (yellow line). Col/Bur and Col/Ws SNP coordinates were generated via a variant calling pipeline described in section 4.2.

In contrast, Col/Ler *msh2-1* crossovers displayed no significant correlation with SNP density (Pearson's r = -0.04; P = 0.434) (fig. 27A,B), whilst Col/CLC *msh2-1* crossovers showed a reduced positive correlation relative to the wild type (Pearson's r = 0.31; $P = 9.04\times10^{-9}$) (fig. 27 C,D). These results indicate that crossovers are redistributing into relatively SNP depleted regions in the *msh2-1* mutant, at the mega-base scale. This implies that MSH2 may be part of a mechanism by which wild type crossovers become positively correlated with interhomolog polymorphism, a phenomenon observed for Arabidopsis hybrids derived from multiple accessions (fig. 27, 30A) (Choi et al., 2013; Ziolkowski et al., 2015; Serra et al., 2018a).

To assess the relationship between crossovers and SNPs at finer scales, I calculated total SNP numbers around each crossover site using varying window sizes, centred on the crossover midpoints assigned by TIGER. As an additional control, four sets of random genomic coordinates were generated – for the Col/Ler and Col/CLC wild type and *msh2-1* mutant – each with the same number of crossovers per chromosome as those occurring in the corresponding genotype. Random coordinates were then adjusted to the midpoint of their 5′ and 3′ SNP coordinates, in an analogous manner to GBS crossover data. At all scales analysed, from 1 to 100 kb, wild type crossovers occurred in regions of elevated SNP density compared to random (fig. 31). These increases were highly significant for both the Col/Ler and Col/CLC wildtypes, at all interval sizes ($P = 6.63\times10^{-9}$–1.56×10^{-43} and 7.43×10^{-25}–1.01×10^{-39}, respectively; Dunn test) (appendix 7.14). In contrast, both the Col/Ler and Col/CLC *msh2-1* crossovers occurred in regions of significantly lower SNP density in comparison to wild type, at all window sizes ($P = 1.38\times10^{-8}$–1.02×10^{-10} and 1.16×10^{-5}–4.90×10^{-7}, respectively; Dunn test) (fig. 31). However, crossovers in the Col/Ler *msh2-1* mutant had significantly higher SNP numbers, relative to their matched random control, for the windows of 1 and 5 kb ($P = 1.02\times10^{-8}$ and 8.78×10^{-3}, respectively; Dunn test), but from 10 kb upwards *msh2-1* crossovers were not significantly different from random (fig. 31A) (appendix 7.14). In contrast, the Col/CLC *msh2-1* mutant crossovers occurred in regions of significantly higher SNP density, at all window sizes (fig. 20B) (appendix 7.14). These comparisons suggest that MMR activity may in fact facilitate, whether directly or indirectly, the formation of crossovers in relatively SNP dense chromosome regions in Arabidopsis.

To assess the distribution of SNP numbers surrounding crossover sites, I visualised these values in a series of histograms (fig. 32). These plots compare SNP numbers around wild type crossovers to *msh2-1* crossovers, or the equivalent matched random datasets, for each window size (fig. 32). Histograms were normalised by the total crossover number in each dataset, to allow for comparisons between wild type and *msh2-1* genotypes. There was considerable overlap between the wild type and *msh2-1* distributions, at all window sizes. However, the *msh2-1* distribution was consistently shifted to the left, as indicated by the blue

bars on the left-hand side of the histograms, indicating that the distribution was biased towards regions of lower diversity. In contrast, the comparisons between the corresponding random datasets revealed no consistent shift in distribution. This analysis confirmed that the reduced SNP density around crossovers in *msh2-1* is caused by a shift in distribution, rather than being driven by a small number of outliers or a change in the shape of the distribution.

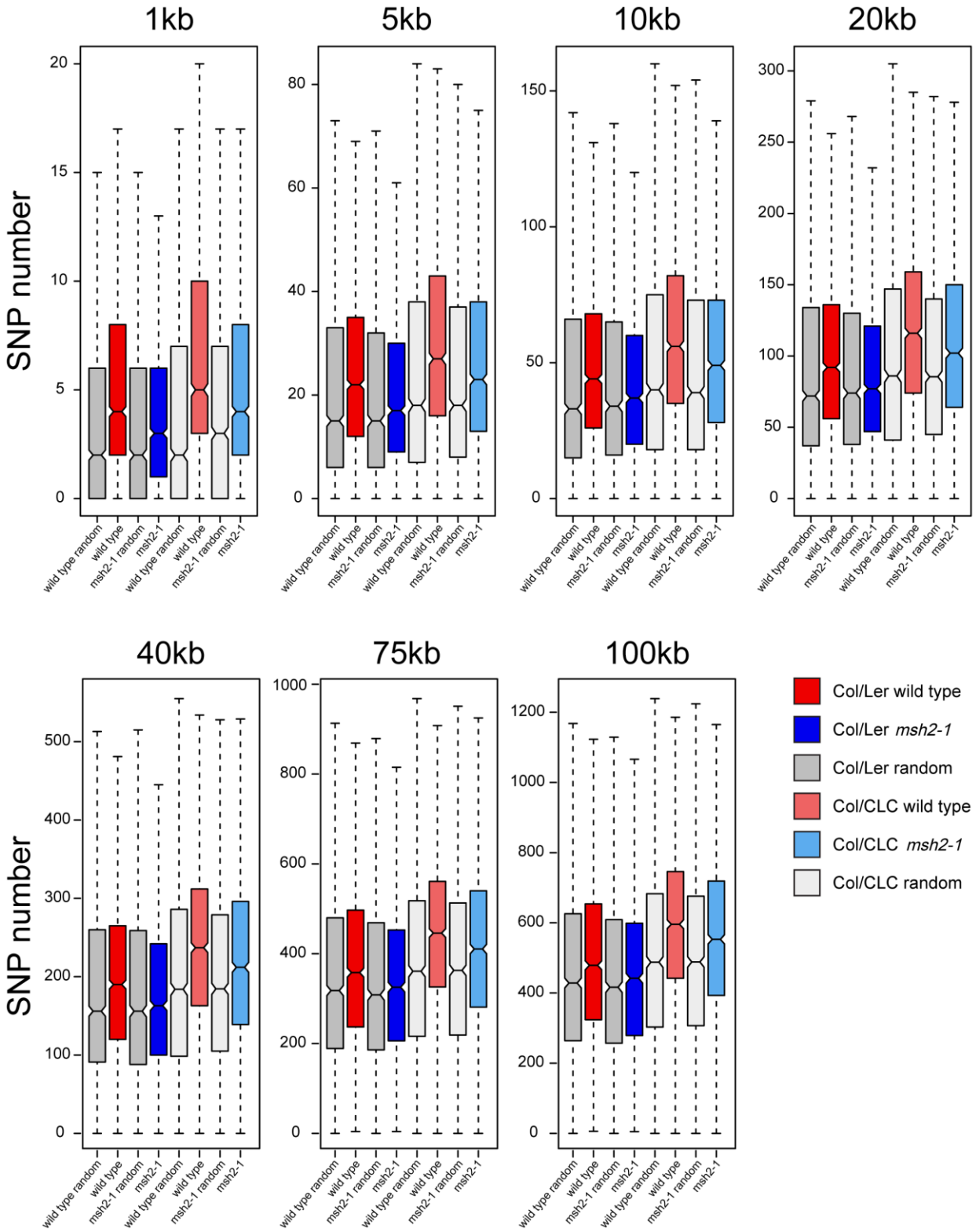

1kb
5kb
10kb
20kb
40kb
75kb
100kb
SNP number
wild type random
wild type
msh2-1 random
msh2-1
Col/Ler wild type
Col/Ler msh2-1
Col/Ler random
Col/CLC wild type
Col/CLC msh2-1
Col/CLC random

Figure 31. SNP density around crossover sites is lower in the *msh2-1* mutant, across multiple scales.

The SNP number surrounding each GBS-defined crossover midpoint was calculated and summed for all crossovers in each genotype, for Col/Ler wild type (red), Col/Ler *msh2-1* (blue), Col/CLC wild type (pale red) and Col/CLC *msh2-1* (pale blue). Each genotype was compared to a corresponding set of random genomic coordinates, adjusted to the midpoint of their 5′ and 3′ SNPs. The procedure was performed using windows of 1, 5, 10, 20, 40, 75 and 100 kb, centred on the crossover midpoints. The data are visualised as box plots, showing the mean, interquartile range and range. *P*-values were calculated using the Dunn test and adjusted for multiple testing with the 'Benjamini–Hochberg' method (appendix 7.14). Col/Ler wild type data are from Serra et al. (2018a). SNP numbers were calculated using the 1,135 genomes datasets (Alonso-Blanco et al., 2016).

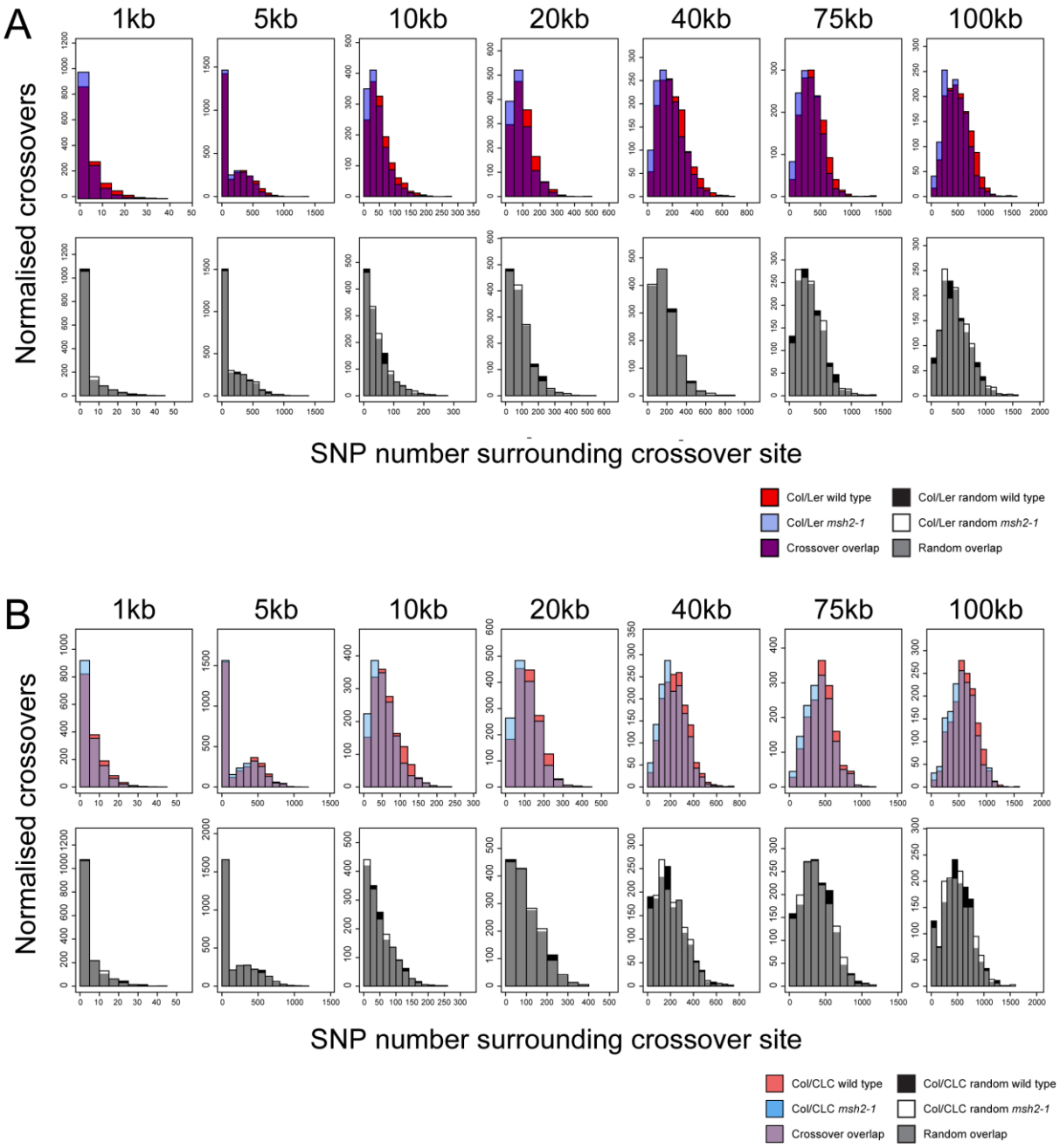


Figure 32. The distribution of SNP number around GBS crossovers sites is altered in the *msh2-1* mutant.

(A) SNP numbers defined for the Col/Ler datasets in fig. 31 are visualised as histograms, showing the number of crossovers at a given SNP number for each genotype. Distributions are shown for windows of 1, 5, 10, 20, 40, 75 and 100 kb, centred on the crossover midpoint. For the upper panel the wild type distribution is shown in red, the *msh2-1* mutant in blue, and the overlap in purple. For the lower panel the corresponding wildtype random distribution is shown in black, the *msh2-1* random in white, and the overlap in grey. (B) As described in (A) but for the Col/CLC hybrid. Col/Ler wild type data are from Serra et al. (2018a). SNP numbers were calculated using the 1,135 genomes datasets (Alonso-Blanco et al., 2016).

4.7 Sites of meiotic crossover are locally associated with high SNP density at the fine-scale

I next investigated the spatial distribution of SNPs surrounding wild type and *msh2-1* crossovers, within 10 kb windows. For this analysis, I counted the presence or absence of SNPs at each genomic coordinate within 5 kb upstream or downstream of each crossover midpoint, and summed these counts over all crossovers for each genotype. This analysis was also performed on corresponding sets of 4 million random crossover coordinates, generated independently for each hybrid background, in order to approximate the genome average expectation. The random coordinates were adjusted to the midpoint of their upstream and downstream SNPs, in a manner analogous to the 'midpoint' positioning of GBS crossover calls. These count vectors were then normalised by the total crossover number for each genotype, to enable comparisons between genotypes. The rolling mean of these values was then plotted for each genotype, using a 300 bp sliding window (fig. 33). It is important to note that whilst this approach allows for a kilobase-scale assessment of SNP density around crossover sites, this spatial resolution is above that of hotspot-based approaches such as pollen-typing or sperm-typing (Cole et al., 2010; Serra et al., 2018b).

The Col/Ler wild type crossovers showed a strong association with SNPs close to their midpoints, which was less pronounced towards the edges of the 10 kb window (fig. 33A). A similar pattern was observed for the Col/CLC wild type, but this association was not as pronounced, and SNP density was higher across the entire window, in comparison to the random (fig. 33B). This fine-scale relationship was additionally confirmed using the Col/Bur and Col/Ws wild type crossover datasets, which again showed a positive association with SNPs close to the crossover midpoints (fig. 33C,D). In contrast, all four matched random datasets showed no positive association with SNPs close to the midpoint coordinates. Indeed, this analysis also made evident the effect of adjusting the random genomic coordinates to the midpoint of the 5′ and 3′ SNP coordinates, which causes a corresponding decrease in SNP density at the centre of the 10 kb window (fig. 33). This effect is also evident for the wild type crossover datasets, which displayed a dip in SNP density at the centre of the window due to the assignment of crossovers to SNP midpoints (fig. 33A,B,C,D). In order to assess variability in the datasets, I plotted the SNP distributions from 50 independently derived random coordinate sets, matched to the number of crossovers per chromosome in Col/Ler wild type (fig. 33E). This indicated that mean SNP density varied over the range of 7.3–8.1 SNPs, with a standard deviation of 0.15 SNPs.

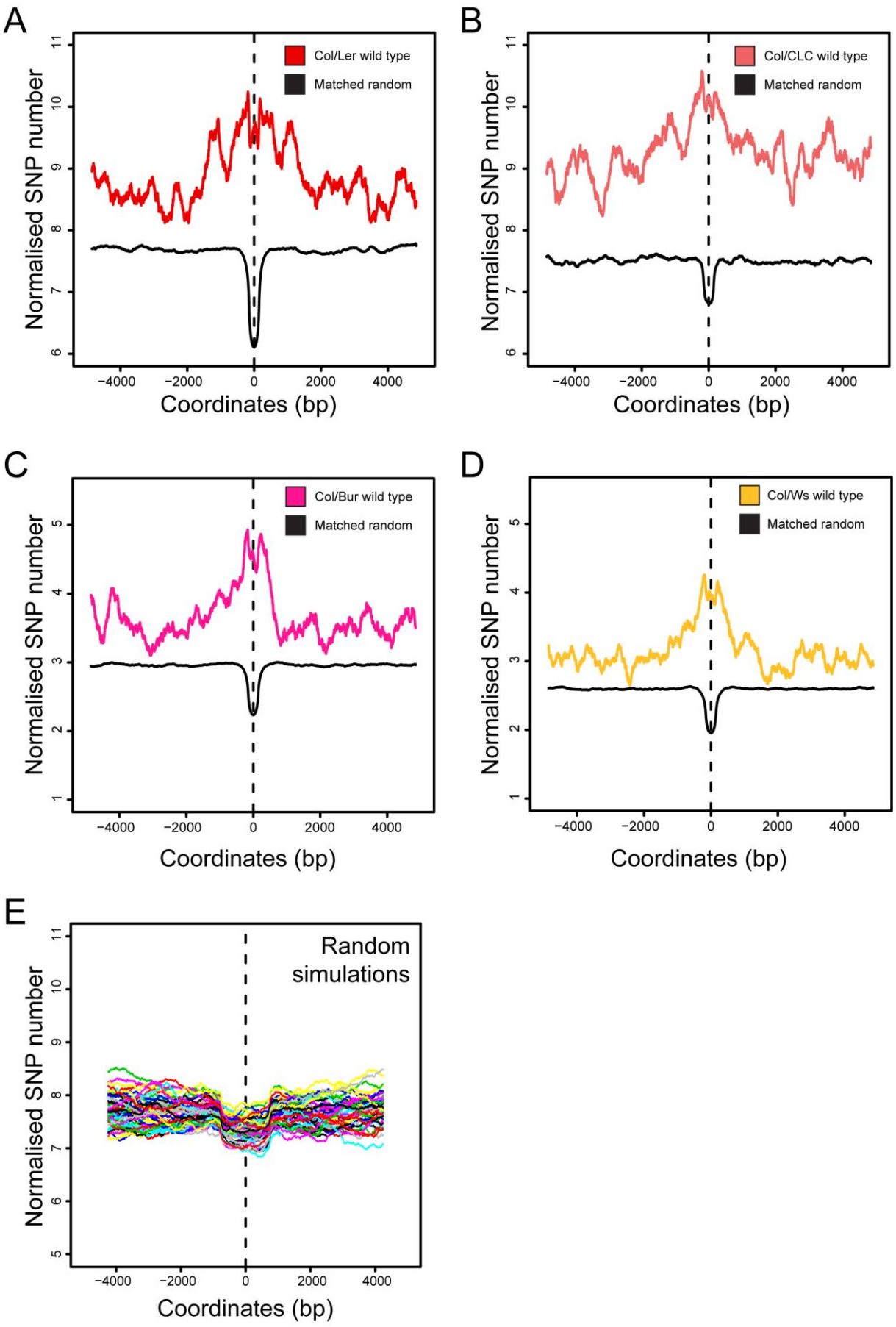

A
Col/Ler wild type
Matched random
Normalised SNP number
Coordinates (bp)
B
Col/CLC wild type
Matched random
Normalised SNP number
Coordinates (bp)
C
Col/Bur wild type
Matched random
Normalised SNP number
Coordinates (bp)
D
Col/Ws wild type
Matched random
Normalised SNP number
Coordinates (bp)
E
Random simulations
Normalised SNP number
Coordinates (bp)

Figure 33. Wild type crossovers are associated with SNPs close to the site of crossover.

SNP distributions were calculated over a 10 kb window, centred on the crossover midpoints. Distributions were calculated using a 300 bp sliding window, and are shown with a set of 4 million random crossovers coordinates (black) for (A) the Col/Ler wild type (red), (B) the Col/CLC wild type (pale red), (C) the Col/Bur wild type (pink), and (D) the Col/Ws wild type (yellow). (E) SNP distributions surrounding crossover sites were calculated for 50 different sets of random genomic coordinates, matched to the number of Col/Ler wild type crossovers, and adjusted to the midpoint of their 3′ and 5′ SNPs. Each simulation is assigned a different colour. The dashed vertical line indicates the centre of the 10 kb window. Col/Ler wild type data are from Serra et al. (2018a). Col/Ler and Col/CLC SNP distributions were calculated using the 1,135 genomes datasets (Alonso-Blanco et al., 2016). Col/Bur and Col/Ws SNP distributions were calculated using variants called from total F_2 sequencing reads, using the pipeline described in section 4.2.

Strikingly, the association between crossover sites and SNP density appeared greatly reduced in the Col/Ler *msh2-1* mutant, which was barely distinguishable from random (fig. 34A). However, there still appeared to be a slight enrichment in SNP density close to the crossover midpoints, relative to the random. This altered relationship was less clear for the Col/CLC *msh2-1* mutant, where total SNP density remained higher than the random across the entire window (fig. 34B). However, the positive association between crossover midpoints and SNP density still appeared reduced in the Col/CLC *msh2-1* mutant, relative to the wild type.

I next investigated SNP distributions across a 10 kb window for crossovers mapped via the same GBS approach in two additional genetic backgrounds. Crossover frequency in the Col/Ler *HEI10* overexpressor is elevated 2-fold genome-wide, via increases in the Class I crossover pathway, and the distribution of crossovers is distalised towards the chromosome ends (Ziolkowski et al., 2017). Using a previously published GBS dataset in this background, I observed a pattern of SNP enrichment close to the midpoint of the 10 kb window, similar to the Col/Ler wild type (Ziolkowski et al., 2017) (fig. 34C). In the *recq4a recq4b* mutant, crossover frequency is increased 3.3-fold and crossovers are highly distalised towards the telomeres (Serra et al., 2018a). Although crossovers in *recq4a recq4b* occurred in regions of lower overall SNP density, they still showed a positive association with SNPs close to the midpoint of the 10 kb window (fig. 34C). These observations are consistent with the general property of Class II crossovers being inhibited by interhomolog polymorphism in Arabidopsis, in contrast to Class I crossovers (Girard et al., 2015; Ziolkowski et al., 2015). Together, these observations reveal an unexpected positive association between meiotic crossovers and regions of local, elevated SNP density observed across multiple genetic backgrounds, and which is paradoxically dependent on *MSH2*.

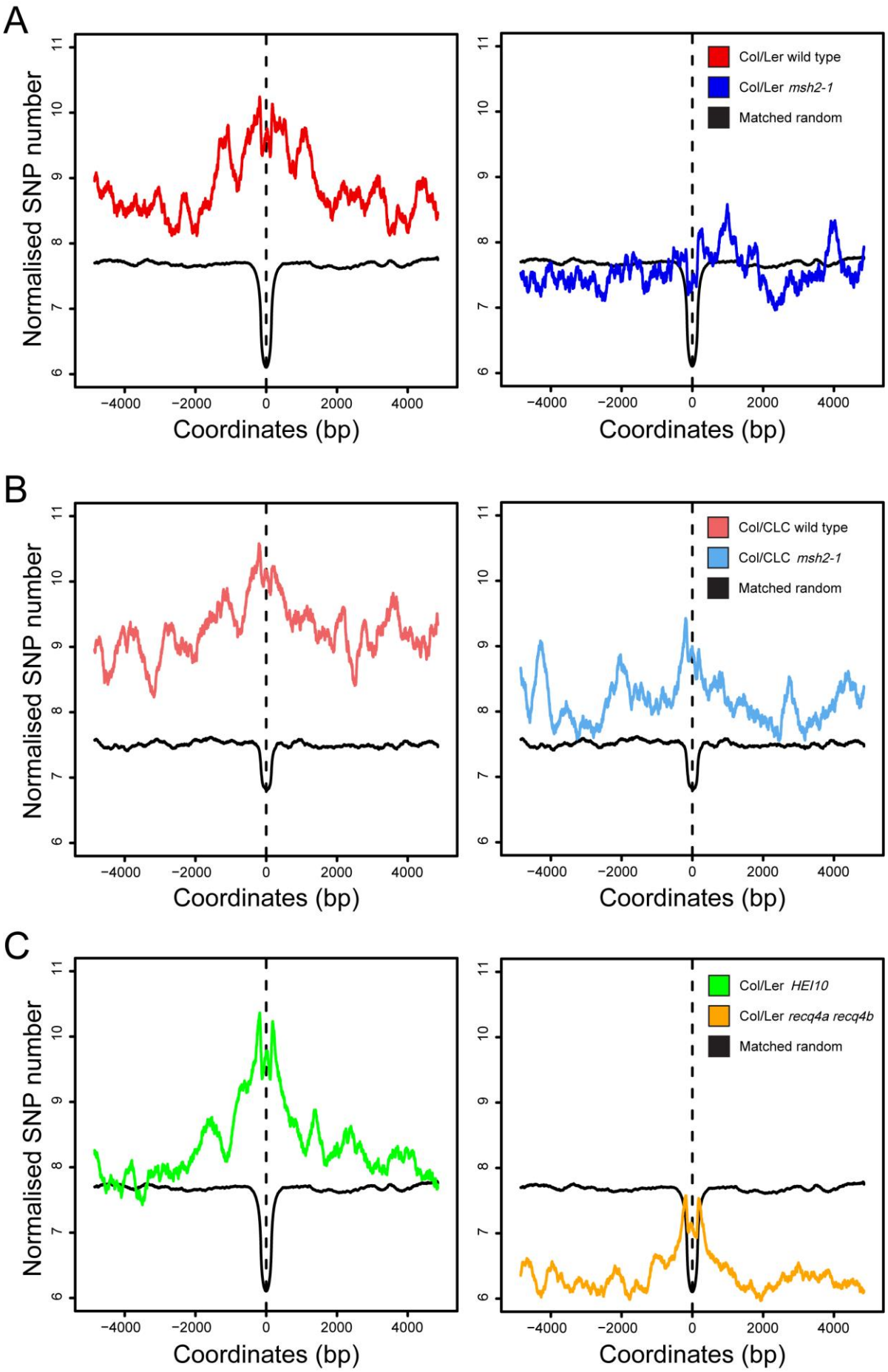

A
Normalised SNP number
Coordinates (bp)
Col/Ler wild type
Col/Ler msh2-1
Matched random
B
Normalised SNP number
Coordinates (bp)
Col/CLC wild type
Col/CLC msh2-1
Matched random
C
Normalised SNP number
Coordinates (bp)
Col/Ler HEI10
Col/Ler recq4a recq4b
Matched random

Figure 34. In contrast to the wild type, crossovers in the *msh2-1* mutant show a reduced association with SNPs close to the site of crossover.

SNP distributions were calculated over a 10 kb window, centred on the crossover midpoints. Distributions were calculated using a 300 bp sliding window, and are shown with a set of 4 million random crossovers coordinates (black) for (A) the Col/Ler wild type (red) and *msh2-1* mutant (blue), (B) the Col/CLC wild type (pale red) and *msh2-1* mutant (pale blue), and (C) the HEI10 overexpressor (green) (Ziolkowski et al., 2017) and the *recq4a recq4b* mutant (orange) (Serra et al., 2018a). The dashed vertical line indicates the centre of the 10 kb window. Col/Ler wild type data are from Serra et al. (2018a). SNP distributions were calculated using the 1,135 genomes datasets (Alonso-Blanco et al., 2016).

A statistical analysis of the differences between each crossover distribution and their matched random dataset broadly confirmed these qualitative trends. The 10 kb window was divided into adjacent 300 bp intervals, and statistical significance was assessed using the Wilcoxon rank sum test. Plotting the $-\log_{10}$ transformed *P*-values with or without correction for multiple testing revealed a highly significant enrichment for SNPs around the crossover midpoints for the Col/Ler and Col/CLC wild types, and the Col/Ler *HEI10* overexpressor (fig. 35). In contrast, although the *msh2-1* mutant hybrids showed a significant association with SNPs at several points across the window, there was no elevated significance close to the centre of the crossover midpoints and no clear relationship between the x-axis coordinates and significance level (fig. 35A,B). A comparison of 25 pairs of random crossover simulations indicated no pattern of significance across the window (fig. 35D). Although many random pairs crossed the significance thresholds at parts of their distribution, there was no clear association between these points and the midpoints of the underlying 10 kb windows. This analysis is not shown for the Col/Ler *recq4a recq4b* mutant dataset, as the SNP density of this mutant was below that of the matched random, resulting in the inverse pattern of significance across the interval. Together, these comparisons provide statistical support for the observed relationships between crossovers and SNPs at the fine-scale, and indicate that losing MMR activity causes a loss of the positive association between crossovers and SNP density at <10 kb scale.

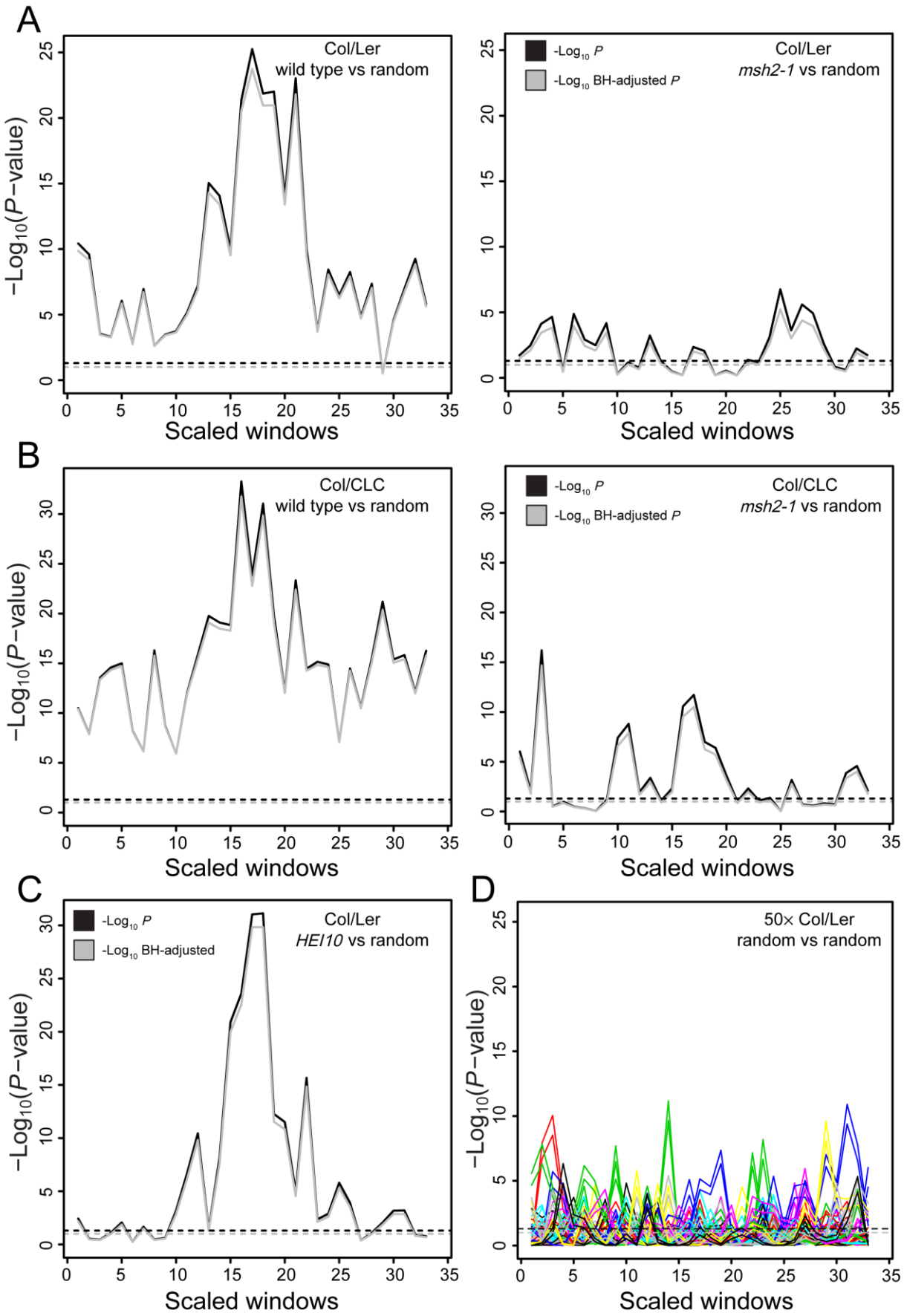

A
Col/Ler
wild type vs random
-Log10 P
-Log10 BH-adjusted P
Col/Ler
msh2-1 vs random
−Log10(P−value)
Scaled windows
B
Col/CLC
wild type vs random
Col/CLC
msh2-1 vs random
C
-Log10 P
-Log10 BH-adjusted
Col/Ler
HEI10 vs random
D
50× Col/Ler
random vs random

Figure 35. SNP density is highly significantly enriched close to wild type crossovers, in contrast to *msh2-1* mutant crossovers.

Statistically significant differences between the 10 kb SNP distributions of each crossover dataset and their corresponding random dataset was assessed across the window. The two distributions were divided into 300 bp adjacent windows, and *P*-values were calculated with Wilcoxon rank sum tests and adjusted using the 'BH' method. These values were $-\log_{10}$ transformed and plotted along a 10 kb axis divided into 34 windows. $-\log_{10}$ transformed *P*-values (black) and $-\log_{10}$ transformed adjusted *P*-values (grey) are plotted for (A) Col/Ler wild type and *msh2-1* mutant, (B) Col/CLC wildtype and *msh2-1* mutant, (C) Col/Ler HEI10 overexpressor (Ziolkowski et al., 2017), and (D) 25 pairs of random crossover coordinates (each pair has an identical colour for adjusted and unadjusted P-values, with the lower line always indicating the adjusted values.) Horizontal dashed lines correspond to an unadjusted *P*-value threshold of $-\log_{10}(0.05)$ (black) and adjusted-*P*-value threshold of $-\log_{10}(0.1)$ (grey). Col/Ler wild type data are from Serra et al. (2018a). SNP numbers were calculated using the 1,135 genomes datasets (Alonso-Blanco et al., 2016).

4.8 The *msh2-1* mutation has no effect on meiotic crossover within large scale structural polymorphisms.

Large scale structural polymorphism has been shown to inhibit meiotic recombination, in a wide range of eukaryotic models (Salomé et al., 2012; Crown et al., 2018; León-Ortiz et al., 2018). I therefore assessed whether large scale structural variation continued to supress crossover formation in the *msh2-1* mutant, using a set of 47 previously reported Col/Ler inversions (Zapata et al., 2016). These inversions comprise a total of 1.59 Mb and have a mean width of 33.8 kb (Zapata et al., 2016), and have been shown to associate with inhibited meiotic crossover (Serra et al., 2018a). In contrast to the relationship to SNPs, the loss of MMR had no noticeable effect on crossover occurrence within the inversion intervals: 1/1,598 crossovers were detected in the Col/Ler wild type and 0/1,309 crossovers in the Col/Ler *msh2-1* mutant ($P = 1.0$; X^2 test), which was significantly fewer than the 30/1,598 crossovers occurring within the inversion intervals for an equivalent number of random coordinates that match Col/Ler wild type crossovers ($P = 5.64\times10^{-7}$ and 2.01×10^{-6}, respectively; X^2 test). Likewise, a large 1.17 Mb inversion on the left arm of chromosome 4 contained no crossovers in either the wild type or *msh2-1* Col/Ler crossover sets, compared to 22 in the random crossover set ($P = 8.18\times10^{-6}$ and 5.90×10^{-5}, respectively; X^2 test), again consistent with previous studies (Salomé et al., 2012; Fransz et al., 2016; Serra et al., 2018a). Together, these observations indicate that loss of MMR activity in Arabidopsis does not facilitate or increase meiotic crossover within large scale inversions.

4.9 H3K4me3 is putatively enriched surrounding *msh2-1* crossovers, in comparison to wild type.

A systematic analysis of histone readers in Arabidopsis revealed that MSH6 contains a Tudor domain, known to bind to H3K4me3 modified nucleosomes (Zhao et al., 2018). Hence, I investigated the potential relationship between crossover positions and nucleosome occupancy or H3K4me3 enrichment, by calculating the level of these features within Col/Ler wild type and *msh2-1* GBS crossover intervals (Serra et al., 2018a; Choi et al., 2018; Lambing et al., 2019). Crossover intervals were defined by the 5' and 3' SNPs used to define genotype changes associated with crossover junctions within F_2 individuals. As the distance between adjacent SNPs varies, MNase-seq and H3K4me3 enrichment were normalised by interval width and plotted across the mean interval width, including 5 kb flanking regions. As a control, enrichment was calculated within a set of random genomic intervals of matched widths. One caveat to this analysis is that crossover positions are derived from Col/Ler F_2 recombinants, whereas the chromatin profiles are derived from Col-0 inbred backgrounds. Hence, these crossover positions may not be representative of crossovers positions in the Col-0 context, due to a combination of *cis* and *trans* effects caused by interhomolog polymorphism. Col/CLC GBS crossovers were not included in this analysis, due to the dramatic alterations to the epigenetic landscape in Cvi-0 (fig. 25B,C) (Kawakatsu et al., 2016; Picard & Gehring, 2017). However, this analysis revealed that whilst Col/Ler wild type and *msh2-1* crossovers occurred at regions with similar nucleosome occupancy, as measured in the Col-0 wild type, *msh2-1* crossovers formed more frequently in regions known to display higher H3K4me3 in the Col-0 wild type (fig. 36). Therefore, this analysis is consistent with an association between MSH2-heterodimers and H3K4me3 (Zhao et al., 2018), suggesting that MMR preferentially suppresses crossovers in proximity to active chromatin in a hybrid context. However, an important caveat to this analysis is that the nucleosome occupancy and H3K4me3 data was derived from Col-0 material, as these datasets do not exit for Arabidopsis *msh2-1* mutants. Hence, the validity of this analysis depends on the loss of MMR having no effect on the genomic enrichment of H3K4me3.

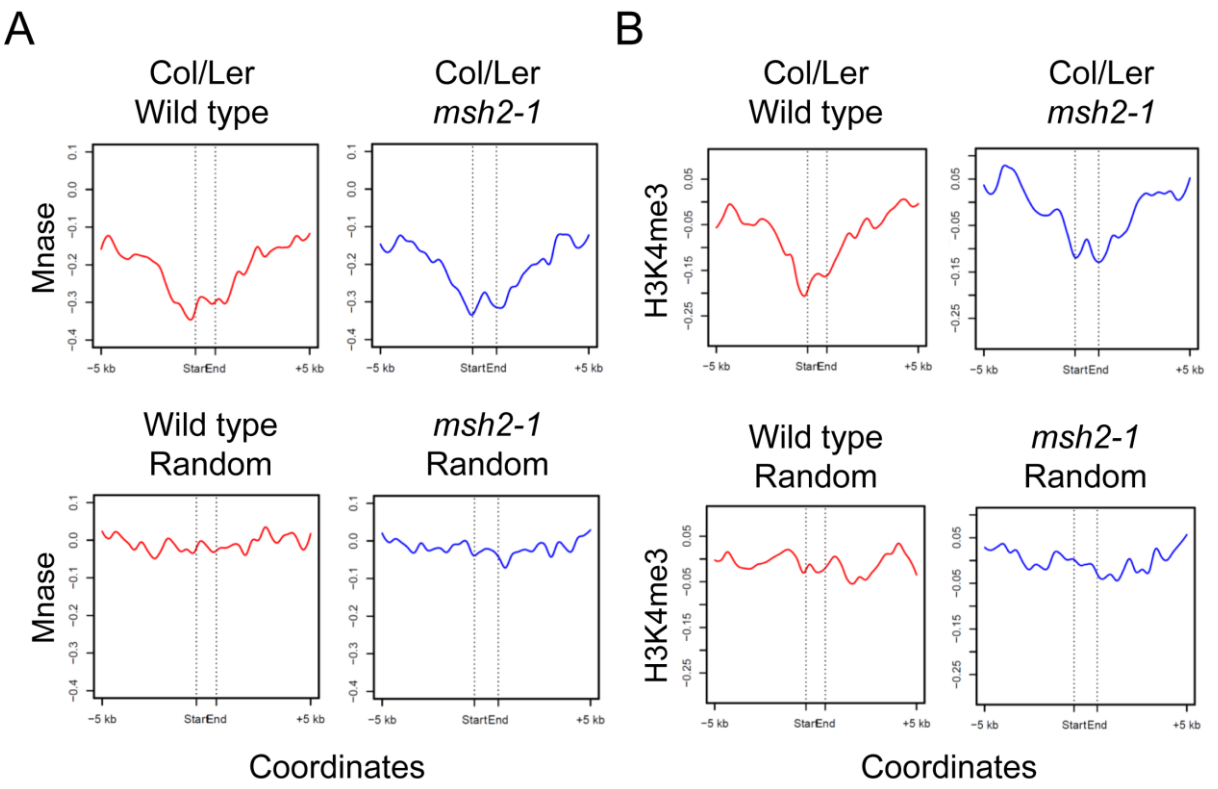


Figure 36. Analysis of nucleosome occupancy and H3K4me3 enrichment surrounding wild type or *msh2-1* crossovers.

(A) Plots of HA-MSH4 $\log_2$(ChIP/input) enrichment (red) within 3,320 GBS crossover intervals, defined by 5' and 3' SNPs demarcating the crossover position, including 5 kb flanking regions (Serra et al., 2018a). (B) As for (A), but showing H3K4me3 enrichment surrounding crossover intervals.

MNase-seq (blue) (Choi et al., 2018; Lambing et al., 2019).

4.10 Discussion

In Arabidopsis intra-species hybrids, meiotic crossovers are positively associated with interhomolog polymorphism at the megabase-scale (Ziolkowski et al., 2015; Fernandes et al., 2018a; Serra et al., 2018a). The results presented here confirm this relationship by experimentally measuring genome-wide crossover landscapes at high resolution in additional Arabidopsis hybrids. With the exception of the Col/Ws wild type hybrid, crossover frequency was positively correlated with interhomolog polymorphism genome-wide. However, given the scale of epigenetic variation across the Arabidopsis genome, which, like polymorphism, shows distinct megabase-scale variation (fig. 25B,C) (Stroud et al., 2013, 2014), it becomes a challenge to disentangle cause and effect relationships due to the correlation of polymorphism with other elements of genome organisation and architecture. The requirement for genetic markers precludes the possibility of measuring the meiotic crossover landscape in an inbred

background. However, it may prove informative to establish a lower-resolution crossover landscape in a near-isogenic Arabidopsis hybrid, such as between a Col-0 wild type and a Col-0 line fixed for mutations caused by MMS mutagenesis. Indeed, crossover landscapes measured in rice Nipponbare × Dongjin F_1 hybrids with a SNP density of 1 per 11 kb found that the *fancm* mutation increased crossover frequency by 2.3-fold, despite these increases being repressed by interhomolog polymorphism in Arabidopsis hybrids, which has higher levels of polymorphism density (e.g. ~1 SNP per 200 bp) (Giraut et al., 2015; Ziolkowski et al., 2015; Mieulet et al., 2018). This result suggests that a SNP density lower than 1 per 10 kb would enable crossover mapping in a system where meiotic recombination may behave as though chromosomes were homozygous. Although it would be impossible to exclude the role that MMS mutations might play in impacting meiotic recombination, crossover mapping using several independent MMS mutant lines would be informative.

In attempting to directly address the role of interhomolog polymorphism in crossover occurrence, by using adjacent megabase-scale genetic intervals with either heterozygous or homozygous sequences, Ziolkowski et al. (2015) observed a redistribution of crossovers into the heterozygous interval, at the expense of crossovers in the homozygous interval (Ziolkowski et al., 2015). This redistribution across juxtaposed heterozygous-homozygous sequences was dependent on the Class I interfering crossover pathway (Ziolkowski et al., 2015). These results suggest that a similar effect may be taking place during mieiotic recombination in Arabidopsis hybrids, which have regional variation in interhomolog polymorphism density at the megabase-scale (fig. 12A) (Alonso-Blanco et al., 2016). In the context of the current data, where loss of MMR activity causes a redistribution of meiotic crossovers, it appears that genome-wide polymorphism density may directly shape the recombination landscape, rather than correlating with other causal features. It is difficult to think how *msh2-1* could affect meiotic recombination if not via mismatches arising following interhomolog strand invasion.

At the chromosome scale, crossovers in the Col/Ler and Col/CLC *msh2-1* mutants redistributed away from the diverse centromere proximal regions, towards the chromosome arms. Relative to wild type, this redistribution reduced the positive correlation between crossovers and interhomolog polymorphism for the Col/CLC hybrid, and abolished the correlation for the Col/Ler hybrid. Curiously, the pattern of crossover redistribution was almost indistinguishable to that seen for the *HEI10* overexpressor, despite this background having a ~2-fold increase in total crossover number (Ziolkowski et al., 2017). In relation to the *cis* effect's dependence on Class I crossovers, this result argues against the genome-wide association between interhomolog polymorphism and crossovers being driven by the same process, as this would predict that increasing Class I crossovers would bias crossover increases towards more polymorphic regions (Ziolkowski et al., 2015). However, the centromere and

pericentromere have also proven intransigent to crossover increases in meiotic mutants with increased class II crossovers, such as *recq4a recq4b* (Serra et al., 2018a). This suggests that the redistribution in *HEI10* may be driven by regional differences in crossover competency. This again highlights the difficulty of disentangling the relationships between sequence divergence and epigenetic variation when crossover patterns are changed along the telomere-centromere axis.

At the finer scale, I also observed a striking enrichment of SNP density in proximity to sites of meiotic crossover. This was most clearly visualised by plotting averaged SNP density at all coordinates across a 10 kb window, averaged for all GBS crossovers from each genotype. In comparison to random control coordinates, also centred on the midpoint of their flanking 5′ and 3′ SNPs, wild type crossovers showed a highly significant association with SNP density at the midpoint of the window. Strikingly, this association was greatly reduced in the *msh2-1* Col/Ler mutant, and to a lesser extent in the *msh2-1* Col/CLC mutant. One concern is that this association is reduced in the *msh2-1* mutant due to the redistribution of crossovers into regions of lower interhomolog polymorphism density, which causes a subsequent increase in the distance between crossover midpoints and their flanking SNPs. However, the *HEI10* overexpressor, which shows a distal increase in crossovers similar to *msh2-1*, retained a clear positive association with SNPs in proximity to sites of meiotic crossover (Ziolkowski et al., 2017; Serra et al., 2018a). This positive association was also observed for *recq4a recq4b*, despite the SNP density across the 10 kb window being lower than for the random coordinates. This comparison argues that the loss of association in the *msh2-1* mutants is not an artefact of crossovers redistributing into regions of reduced polymorphism at the chromosome scale. These findings prove challenging to unify with the observation that local interhomolog polymorphism represses crossover frequency (see section 6.3 for further discussion) (Serra et al., 2018b).

One technical concern, in relation to the fine-scale association between SNPs and crossover positions, relates to the accuracy of the GBS methodology to resolve crossovers (Rowan et al., 2015). Comparing the same seven GBS libraries sequenced at low or high depth (mean coverage = 3.9 or 18.1, respectively) revealed that total crossover number differed in only a single F_2 individual (low = 7, high = 6) (Serra et al., 2018a). However, a reanalysis of the crossover midpoints derived from high or low sequencing revealed that only 13% were called with an identical SNP interval midpoint. The median difference between the low and high sequencing crossover sites was 1,137 bp, which is consistent with estimates of the accuracy of GBS crossovers based on Sanger sequencing confirmation of putative crossover junctions (Rowan et al., 2015). This previous analysis confirmed eight of eleven crossovers, and found that almost all occurred within 1.5 kb of the predicted flanking markers called with the GBS

analysis (Rowan et al., 2015). However, a closer inspection of the low/high GBS crossover sites revealed that 48% were within 1 kb of each other, and 78% were within 5 kb of each other (Serra et al., 2018a). However, three sites differed by 100 kb, 111 kb and 2.3 Mb, which indicates that it is possible for TIGER predictions to very greatly (Serra et al., 2018a). This indicates that caution must be taken when interpreting the location of GBS crossovers at the fine-scale.

A further consideration in relation to the fine-scale relationships is how marker density may influence the probability of crossover assignment. TIGER assesses the support for each marker based on a sliding window of 10,000 markers, which are assessed independently of physical distance between the markers. This raises the concern that this distance-independent approach may be biasing crossover calls towards regions of local SNP density, due to the increased statistical support in these regions. It will be important to perform a rigorous assessment of this procedure to identify any possible biases. It will also be useful to incorporate the assigning of confidence estimates to GBS crossover sites, and to assess any differences between low and high confidence crossovers.

However, even given these concerns, the observation that the *msh2-1* mutation alters the fine-scale distribution of SNP density surrounding crossover sites, relative to other crossover mutants, strongly indicates that this relationship is *bona fide*. A priority will be to confirm this fine-scale relationship using additional TIGER-independent methodologies. One promising candidate is a recently developed method for assessing genome-wide crossover landscapes that utilises linked-read sequencing technology (Zheng et al., 2016). This method produces short reads derived from the same (~45 kb) DNA fragments, hence improving the confidence of mapping, and assigns crossovers based on the Long Ranger pipeline (Zheng et al., 2016). Crucially, this analysis approach utilises reads that contain variants corresponding to both parental alleles, as these reads are most likely to span the crossover positon. This approach has been used to map crossover landscapes in sticklebacks, mouse and Arabidopsis, and will be an important tool for confirming results from GBS analyses (Dréau et al., 2018; Sun et al., 2018).

An important experimental consideration is how the level of interhomolog polymorphism used in this study relates to our knowledge of recombination mechanisms and intermediates. In Arabidopsis, crossovers-associated gene conversion tracts are approximately ~400 bp, whilst noncrossover conversion tracts are estimated at less than 50 bp (Drouaud et al., 2013; Wijnker et al., 2013). For the Arabidopsis hybrids used in this study, the average interhomolog polymorphism density ranged from 1 SNP/indel per 209 bp in the Col/Ler hybrid to 1 SNP/indel per 183 bp in the Col/CLC hybrids. However, given the extensive variation in polymorphism

density across the genome, a significant proportion of the genome will be either above or below these levels. Hence, it seems likely that crossover formation will be impacted by polymorphism to a greater extent than noncrossovers, given the greater tract length associated with these events (Drouaud et al., 2013; Wijnker et al., 2013). However, it is also likely that many recombination events in these hybrids will not experience a mismatch. This effect may explain the lack of either increased chromosome nondisjunction or fertility defects in the Arabidopsis hybrids, and the lack of a significant increase in total crossover number in the *msh2-1* mutant.

Although this investigation was not designed to assess the possible role of MMR in repressing meiotic recombination within large-scale polymorphisms, I was able to compare the GBS crossover positions to a series of large-scale polymorphisms identified in a high quality L*er*-0 genome assembly (Zapata et al., 2016). Large-scale polymorphism is caused by genome rearrangements in the L*er*-0 genome, relative to Col-0, such as inversions or duplications. In particular, the L*er*-0 genome contains a large 1.17 Mb inversion on the short arm of chromosome 4, previously characterised as a potent repressor of meiotic crossover (Salomé et al., 2012; Fransz et al., 2016; Serra et al., 2018a). In the *msh2-1* mutant, crossover frequency was repressed and indistinguishable from the wild type. This observation is consistent with the hypothesis of Goldman & Lichten (2000) that ectopic recombination between non-allelic sequences would be repressed due to the restriction of interhomolog interactions between these regions, rather than the action of MMR. Furthermore, this data is also consistent with a recent study of meiotic recombination in MMR-deficient *C. elegans*, where meiotic crossovers within an 8 Mb inversion heterozygote did not increase in either the *msh2*, *msh6*, *mlh1* or *pms2* mutant backgrounds (Leon-Ortiz et al., 2018). Curiously, the crossover increases in the *C. elegans* anti-crossover helicase RTEL-1 mutant, observed within both the inversion and two additional syntenic genetic intervals, were repressed by mutating *MSH2* (Leon-Ortiz et al., 2018). It would therefore be interesting to assess the genome-wide crossover increases in an analogous Arabidopsis genetic background, such as the *recq4a rec4b msh2-1* triple mutant.

The central challenge presented by the current study becomes how to integrate these findings with studies of MMR-activity in regulating meiotic recombination in other experimental systems, and to contextualise these observations into the extensive literature on both the molecular mechanisms of MMR and recombination. How these observations can be reconciled with measurements of genome-wide meiotic recombination in additional MMR-deficient eukaryotic models, particularly budding yeast (Cooper et al., 2018), as well as how these can be interpreted at a mechanistic level, will be the focus of chapter 6.

4.11 Acknowledgements

I would like to thank Ms Pallas Kuo for performing heterochromatin quantification of pachytene spreads, Dr Christophe Lambing for advice on the GBS protocol, and Dr Andy Tock for help with bioinformatic and statistical analyses. I would also like to thank both Dr Christophe Lambing and Dr Emma Lawrence for sharing unpublished crossover datasets.

Chapter Five – Results – Investigating the genome-wide distribution of MSH4.

5.1 Introduction

In addition to the role of eukaryotic MutS homolog (MSH) proteins in detecting mismatched basepairs, two members of the MSH family play a specialised role in regulating meiotic crossovers (Ross-Macdonald & Roeder, 1994; Hollingsworth et al., 1995; Novak et al., 2001). MSH4 and MSH5 play no known role in mitotic MMR, and instead form a meiosis-specific heterodimer (MutSγ) (Bocker et al., 1999; Paquis-Flucklinger et al., 1997; Snowden et al., 2004). Both MSH4 and MSH5 are members of the conserved ZMM pathway, and are required for Class I interfering crossovers in plants (Higgins et al., 2004, 2008b). The *msh4* mutant phenocopies *msh5* in Arabidopsis, with no cytological evidence for MSH5 loading onto chromatin in *msh4*, suggesting an intact MutSγ complex is required for recruitment to meiotic chromatin (Higgins et al., 2008b; Lu et al., 2008). Loss of MutSγ activity in budding yeast, mouse and Arabidopsis causes loss of crossovers, unbalanced chromosome segregation and decreased spore/gamete viability (Ross-Macdonald & Roeder, 1994; Kneitz et al., 2000; Higgins et al., 2008b). This meiotic role is deeply conserved in eukaryotes, and MutSγ is required for wild type levels of crossover in budding yeast, mouse and Arabidopsis (Ross-Macdonald & Roeder, 1994; Hollingsworth et al., 1995; Novak et al., 2001; Kneitz et al., 2000; Higgins et al., 2008b).

Compared to MSH2, MSH3, MSH6 and MSH7, both MSH4 and MSH5 have lost their N-terminal domain I, which is known to be essential for mismatch recognition in MutS homologs (Ross-Macdonald & Roeder, 1994; Hollingsworth et al., 1995). Structural modelling of the MutSγ complex in budding yeast predicts a greater central cavity between the two subunits, potentially wide enough to enclose larger, more complex DNA substrates than mitotic MutS heterodimers (Rakshambikai et al., 2013).

Both direct biochemical analyses and computational modelling support a model where MutSγ binds to meiotic crossover intermediates, stabilising SEIs and promoting their maturation into dHJs (Snowden et al., 2004, 2008; Rakshambikai et al., 2013; Lahiri et al., 2018). However, *in vitro* analyses indicate that both human and yeast MutSγ have the greatest affinity for HJs, whilst showing almost no affinity for dsDNA, G/T mismatched dsDNA or open junction substrates (Snowden et al., 2004, 2008; Lahiri et al., 2018). In addition, the budding yeast MutSγ was shown to have moderate affinity for 3'-overhangs, ssDNA forks, and displacement

loops (Lahiri et al., 2018). Hence, the *in vivo* target for MutSγ remains unclear, but is likely to be a specific form of DNA joint molecule.

The distribution of MutSγ has been studied during Arabidopsis and rice meiosis. During meiotic prophase, Arabidopsis MutSγ initially forms numerous (~80-100) punctate foci at mid-leptotene, which are associated with the meiotic axis (Higgins et al., 2004). MSH4 foci gradually decline in number through prophase I, and a subset of those foci remaining at early-pachytene transiently co-localise with MLH1, a marker of class I crossovers (Higgins et al., 2008b; Lu et al., 2008). Likewise, both the rice *msh4* and *msh5* mutants have a chiasmata frequency ~10% of the wild type level, leading to chromosome non-disjunction and sterility (Luo et al., 2013; Zhang et al., 2014), although synaptonemal complex assembly is undisturbed (Zhang et al., 2014). Arabidopsis MSH4 foci briefly co-localise with RAD51, and RAD51 loading slightly precedes MSH4 binding (Higgins et al., 2004). Consistent with this pattern of co-localisation, the rice MSH5 protein directly interacts with four RPA homologs, as demonstrated by yeast two-hybrid assays and pull-down assays (Wang et al., 2016). The observation that rice MSH5 can still be loaded onto meiotic chromosomes in the absence of ZIP4, MER3 or HEI10 – whilst those factors cannot localise efficiently in the absence of MSH5 – provides additional evidence for MutSγ acting upstream in the ZMM pathway, potentially shortly after strand invasion (Luo et al., 2013). A similar relationship was observed for the loading of ZIP4 and MER3 in the *msh4* mutant in rice (Zhang et al., 2014). Moreover, the reduction in chiasmata frequency was most pronounced in the rice *msh4* and *msh5* single or double mutants, in comparison to other ZMM mutants such as *zip4*, *mer3* or *hei10* (Shen et al., 2012; Wang et al., 2012; Zhang et al., 2014). This again suggests an epistatic relationship between MutSγ and other proteins in the ZMM pathway. Taken together, these studies support a conserved role for MutSγ early in plant meiosis, where it interacts with and stabilises a subset of early recombination interactions between homologous chromosomes, but then remains present on meiotic chromatin until later stages associated with crossover formation.

Studies in mouse and budding yeast have shown that MutSγ foci number and distribution are regulated by posttranslational modifications (Reynolds et al., 2013; Qiao et al., 2014; Ahuja et al., 2017; Rao et al., 2017; He et al., 2018). Notably, the budding yeast E3 ligase Zip3 recruits the proteasome core complex to meiotic chromosomes during prophase I (Ahuja et al., 2017). This is consistent with work in mice, where antagonistic roles for the ubiquitin ligase HEI10 and the SUMO ligase RNF212 have been defined to stabilise a subset of MutSγ foci and regulate the proteolytic turnover of this complex (Reynolds et al., 2013; Qiao et al., 2014). A Zip3 orthologue, termed HEI10, also exists in plants such as Arabidopsis and rice (Chelysheva et al., 2012; Wang et al., 2012), which is a dosage-sensitive regulator of crossover frequency genome-wide (Ziolkowski et al., 2017), although to date no *RNF212* homolog has been

identified in plants. In rice, a direct physical interaction between MutSγ and HEI10 was observed using yeast two-hybrid and pull-down assays (Zhang et al., 2014), consistent with direct regulation of MutSγ by HEI10 family E3 ligases. Together, these observations suggest that MSH4 is an important regulatory node in the designation and formation of meiotic crossovers in the class I interfering pathway, and is a target of posttranslational regulation by the ubiquitin and/or SUMO pathways.

In plants, our primary understanding of the role of MutSγ in meiosis derives from the use of cytogenetic, molecular genetic and biochemical approaches (Lambing & Heckmann et al., 2018). However, how recombination proteins are distributed across the genome in relation to underlying chromatin features remains poorly understood, particularly for the ZMM family of pro-crossover proteins. Chromatin Immunoprecipitation–sequencing (ChIP-seq) enables the resolution of DNA-associated proteins with a resolution of ~200-400 basepairs, dependent on the method of chromatin fragmentation (Kaufmann et al., 2010). Briefly, tissue is cross-linked with formaldehyde, in order to covalently bind DNA–protein and protein–protein interactions, and nuclei are isolated. Chromatin is then extracted and fragmented, and an antibody is used to pull-down and enrich target epitopes. Cross-linking is reversed, and the DNA associated with the target epitope is isolated and sequenced in order to map the binding positions of the protein of interest. This technique has been successfully performed on meiotic proteins in budding yeast and mouse (Prieler et al., 2005; Kugou et al., 2009; Diagouraga et al., 2018), and in a modified form has been applied to meiotic recombinases in mouse and maize (Smagulova et al., 2011; He et al., 2017).

Adapting ChIP-seq methods to the study of Arabidopsis meiotic proteins is challenging, given the difficulty in isolating large number of meiocytes from floral bud tissues. In collaboration with Dr Christophe Lambing and Dr Kyuha Choi, I contributed to optimising a ChIP-seq protocol for the analysis of meiotically expressed proteins in Arabidopsis (Lambing et al., 2019). This method was adapted for immunoprecipitates meiotic proteins from stage 9 floral buds, which correspond predominantly to male meiosis (Smyth et al., 1990). I worked directly with Dr Christophe Lambing to adapt a ChIP-seq protocol established by Dr Kyuha Choi for histone and SPO11-1 ChIP from Arabidopsis bud tissue (Choi et al., 2018; Lambing et al., 2019). For this approach, we utilised a tissue collection strategy established by Dr Kyuha Choi for the immunoprecipitation of SPO11-1 oligonucleotides from Arabidopsis (Choi et al., 2018). With the guidance of Dr Kyuha Choi and Dr Christophe Lambing I established western blot and immunoprecipitation protocols for MSH4, performed time-course experiments to establish an optimal DNA fragmentation approach, and fine-tuned the ChIP protocol to reduce the quantity of background DNA precipitating with the beads. Following these steps, I performed a qPCR analysis of characterised meiotic hot spots and a scaled up MSH4 ChIP-seq experiment.

Using this approach, I sought to investigate the genome-wide distribution of MutSγ, via ChIP-seq of MSH4. However, the ChIP-seq approach and the results presented in this chapter have several limitations and caveats that must be stated at this point (for an extended discussion of these factors see section 5.14). For instance, the data presented are the results of a single ChIP-seq replicate, and therefore need to be validated by further replicates. Furthermore, due to the small amount of DNA precipitated from the control Col-0 ChIP experiment performed in parallel a control DNA library was not sequenced. Therefore, the results presented here are only normalised by the input chromatin DNA library that was sequenced prior to immunoprecipitation, and therefore the background signal from the α-HA immunoprecipitation step remains unknown. This presents a concern for the data interpretation, as it remains unclear what proportion of the enrichment signal is specific to MSH4 (rather than simply background from the antibody). Additionally, due to the greater number of male gametes (pollen) compared to female gametes in Arabidopsis flowers, this approach enriches for male over female meiocytes which potentially presents a male-biased pattern of ChIP-seq enrichment (Ma et al., 2006; Lambing et al., 2019). As such, this chapter represents the results of a preliminary data analysis of MSH4 ChIP-seq, with a focus on the analysis of the relationship between MSH4 enrichment and the underlying chromatin and recombination landscapes, and will require several additional experiments and controls prior to publication.

5.2 Assembly of tagged *MSH4* constructs

The success of ChIP-seq experiments depends on the specificity and affinity between the chosen antibody and its target epitope, and its ability to recognise the epitope in the context of cross-linked chromatin (Kugou & Ohta, 2009). One approach is to develop stable transgenic lines with a functional epitope-tagged protein, for which there are high affinity monoclonal and polyclonal antibodies that are commercially available. However, this strategy risks interfering with the properties of the endogenous protein, and hence failing to complement the mutant phenotype. However, several studies have previously epitope-tagged MSH4 and reported successful complementation of mutant phenotypes in these species. For example, MSH4 was functionally tagged at its C-terminus with a hemagglutinin (HA) tag in budding yeast (Novak et al., 2001), and with a C-terminal GFP in *Sordaria macrospora* (Storlazzi et al., 2010). Therefore, I adopted a transgenic approach to introduce a tagged version of *MSH4* into the *msh4-1* mutant, in order to produce a translationally fused MSH4-epitope protein suitable for ChIP.

I first assembled binary vectors containing a linker, Myc or HA tag, and a *Nos* terminator (*Nos*T). A linker-5×Myc-*Nos*T sequence was amplified from a previously reported construct that successfully complemented the Arabidopsis *spo11-1-3* mutant using primers ‘LMN_F’ and

‘LMN_R’ (appendix 7.2) (Choi et al., 2018; Underwood et al., 2018). This amplicon was inserted into the multiple cloning site (MCS) of the binary vector *pGreenII-0229*, via restriction cloning at the *Pst*I and *Bam*HI sites (Hellens et al., 2000). To create the second vector, I excised the linker-Myc region with *Pst*I and *Kpn*I, and inserted a linker-3×HA sequence formed by annealing a series of oligonucleotides (‘HA_tag_A’ to ‘HA_tag_G’) (appendix 7.2). In both cases, the linker contained a (GGGS)×4 amino acid repeat sequence, which is designed to improve the accessibility of the tag to its cognate antibody. To assemble C-terminal tagged constructs, the *MSH4* genomic locus, including 1744 bp upstream of the start codon, was PCR amplified from Col-0 genomic DNA as three ~2.4 kb amplicons, with primers ‘MSH4_Ct_1(2,3)F’ and ‘MSH4_Ct_1(2,3)R’. These sequences were inserted via Gibson assembly cloning immediately upstream of the linker-tag-*Nos*T sequence for both Myc and HA vectors, after linearising the vector with *Eco*RV (appendix 7.2) (Gibson et al., 2009). To assemble N-terminal tagged constructs, a 1744 bp promoter and 5'-UTR region of *MSH4* was amplified with primers ‘MSH4_Ct_1F’ and ‘MSH4_prom_R’, and inserted upstream of the linker-tag-*Nos*T, after linearising the vector with *Eco*RV. The *MSH4* gene body was PCR amplified as two ~2.7 kb amplicons with primers ‘MSH4_GB_Myc_F1(2)’ and ‘MSH4_GB_Myc_R1(2)’ or ‘MSH4_GB_HA_F1(2)’ and ‘MSH4_HA_R1(2)’ for the Myc and HA constructs, respectively (appendix 7.2). These were inserted between the tag and *Nos*T using Gibson assembly, after linearising the vector with *Sma*I. Successful cloning was confirmed by restriction digest analysis and Sanger sequencing.

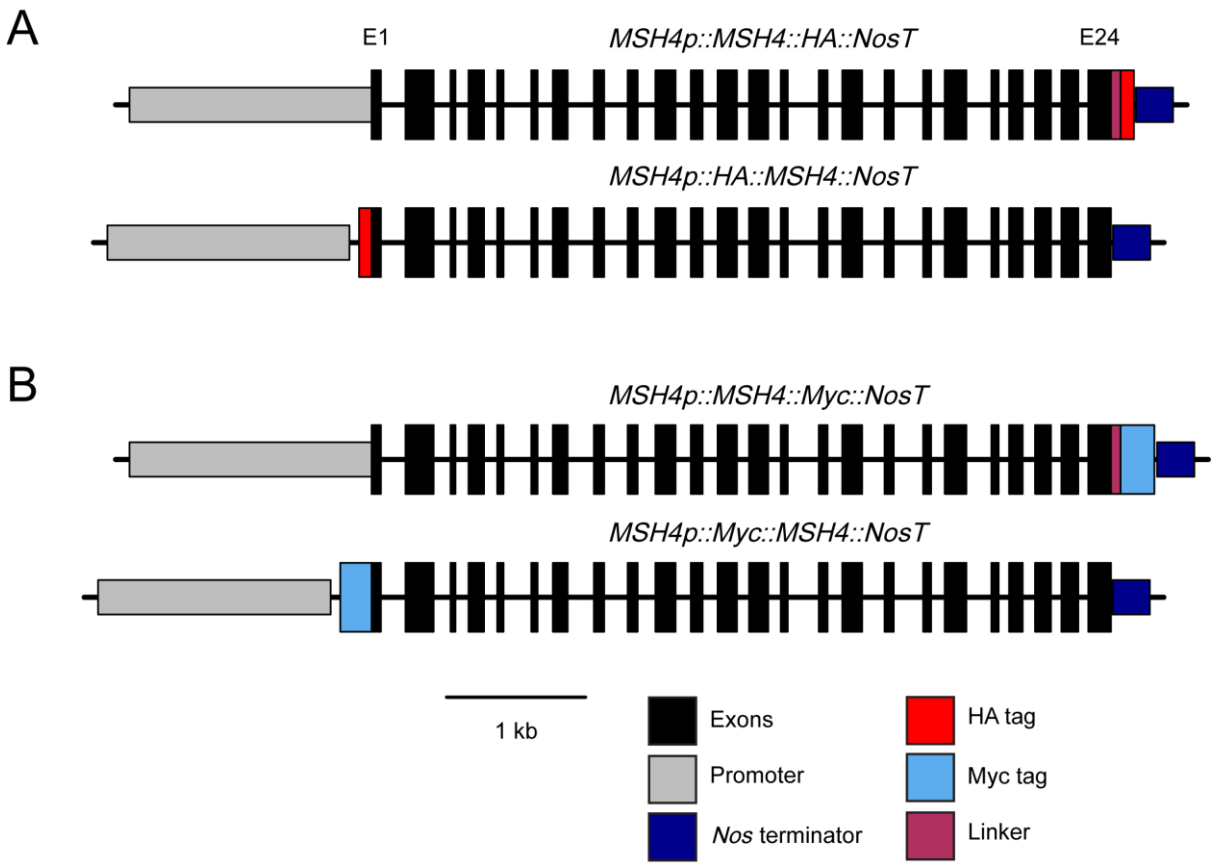


Figure 37. Diagram showing the structure of tagged *MSH4* constructs.

(A) Schematic representation of the C- and N-terminal HA-tagged *MSH4* constructs. Exons (black), promoters and 5'-UTR regions (grey), HA tag (red), linker (purple), and N*os*T (dark blue) are indicated. Introns and non-coding sequence intervening the sequence features are indicated with a horizontal black line. The scale bar = 1 kb. (B) As described for (A) but for the Myc tagged constructs. The Myc tag is indicated in pale blue.

5.3 Transformation of epitope tagged *MSH4* constructs complements the *msh4-1* phenotype.

The four constructs – *MSH4-HA*, *HA-MSH4*, *MSH4-Myc* and *Myc-MSH4* – were then transformed into *msh4-1*/+ heterozygous plants using the *Agrobacterium tumefaciens* strain GV3101, and transformants were selected on phosphinothricin (PPT) supplemented MS medium (Zhang et al., 2006). The presence of a stably integrated vector conferred survival due to the presence of a PPT-resistance gene in *pGreenII-0229*, and was confirmed by PCR genotyping using primers 'T1_HA-MSH4_F' and 'SALK_136296_RP' (appendix 7.2).

Transformants were also genotyped for the *msh4-1* T-DNA or MSH4 wild type locus (appendix 7.2), and homozygous *msh4-1* mutants were assessed for complementation of the mutant phenotype. At the morphological level, the siliques appeared as wild type for all four constructs, indicating successful complementation of the mutant phenotype (fig. 38A). To confirm the presence of the tagged protein, total protein was extracted from meiotic-stage flower buds and a western blot was performed. This failed to reveal the presence of the Myc-tagged MSH4 protein for either the N- or C-terminal constructs (data not shown), despite the constructs successful complementation of the *msh4-1* fertility phenotype (fig. 38A). However, two independent MSH4 lines with a HA tag at their N-terminus showed a band of the expected size (~92 kDa), which was absent from both the Col-0 control and protein extracted from transgenic tagged line leaf tissue (fig. 40A). The absence of a signal of the expected size from *HA-MSH4* transgenic leaf tissue was expected due to the meiosis specific expression of *MSH4* (Higgins et al., 2004). Line 2 (hereafter *HA-MSH4*) showed the strongest signal and was selected for subsequent analysis.

A


wild type
msh4-1
HA-MSH4
msh4-1
MSH4-Myc
msh4-1

B

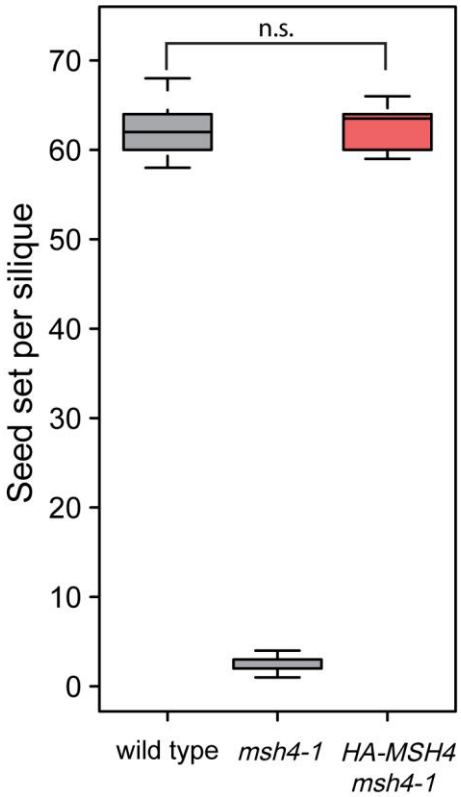

n.s.
Seed set per silique
70
60
50
40
30
20
10
0
wild type
msh4-1
HA-MSH4
msh4-1

Figure 38. *HA-MSH4* complements the *msh4-1* infertility phenotype.

(A) Images of inflorescences from wild type, *msh4-1*, *HA-MSH4 msh4-1*, and *MSH4-Myc msh4-1* plants. Shortened siliques indicate reduced fertility. (B) Box plots indicating seeds per silique for wild type, *msh4-1*, and *HA-MSH4 msh4-1* backgrounds. Box plots indicate median, interquartile range and range. Statistical significance was assessed using the Wilcoxon rank sum test.

To further confirm complementation of the *msh4-1* mutation, seeds per silique were counted in the T_2 population of the *HA-MSH4* line (appendix 7.15). Seed set in the *HA-MSH4 msh4-1* background was not significantly different from wild type ($P = 0.099$; Wilcoxon rank sum test), in contrast to the dramatic reduction in seed set in the *msh4-1* mutant ($P = 2.32\times10^{-11}$; Wilcoxon rank sum test) (fig. 39B). Crossover frequency was analysed at the sub-telomeric *I2f* and *I3c* FTL intervals (0.67 and 1.19 Mb, respectively), and the sub-telomeric *420* seed-based interval (5.11 Mb) (fig. 39A). For all three genetic intervals there was no significant difference between wild type and *HA-MSH4 msh4-1* ($P = 0.629$, 0.686 and 0.223, respectively; Mann-Whitney *U* tests) (fig. 39A) (appendix 7.16 & 7.17). A previously reported transgene overexpressing *HEI10*, which causes a 2-fold elevation in crossover frequency, was also introduced into the *HA-MSH4 msh4-1* background (Ziolkowski et al., 2017), and the increased recombination rate was confirmed using the *420* interval ($P = 0.0005$; Mann-Whitney *U* test) (fig. 39A) (appendix 7.17). Hence, the *HA-MSH4* transgene appears to fully complement the *msh4-1* mutation.

5.4 The presence of HA-MSH4 was validated by immunoprecipitation and western blotting after nuclear isolation.

The previous western blot results were validated using two further approaches. Firstly, HA-MSH4 was immunoprecipitated from Arabidopsis floral buds, after protein extraction and denaturation. Western blotting demonstrated a clear enrichment of the protein at the expected band size of ~92 kDa following immunoprecipitation (IP) (fig. 40B). The band was absent from the Col-0 control and was not detected in an equivalent volume of HA-MSH4 IP input sample, due to the protein being too diluted in comparison to the IP sample (fig. 40B). Notably, the lower, non-specific band observed by western blotting (fig. 40A), was absent after IP (fig. 40B). Secondly, protein was extracted after nuclear isolation from *HA-MSH4* or Col-0 buds, with or without cross-linking, by boiling for 10 minutes in 50 µl of SDS-PAGE sample buffer. For both cross-linked and non-cross-linked samples, the expected band of ~92 kDa was only observed in the *HA-MSH4* samples (fig. 40C). The non-specific band was again absent from the western blot, indicating that the antibody was likely recognising a non-nuclear protein. Together, these results suggested that the *HA-MSH4* line would be amenable to ChIP-seq investigation.

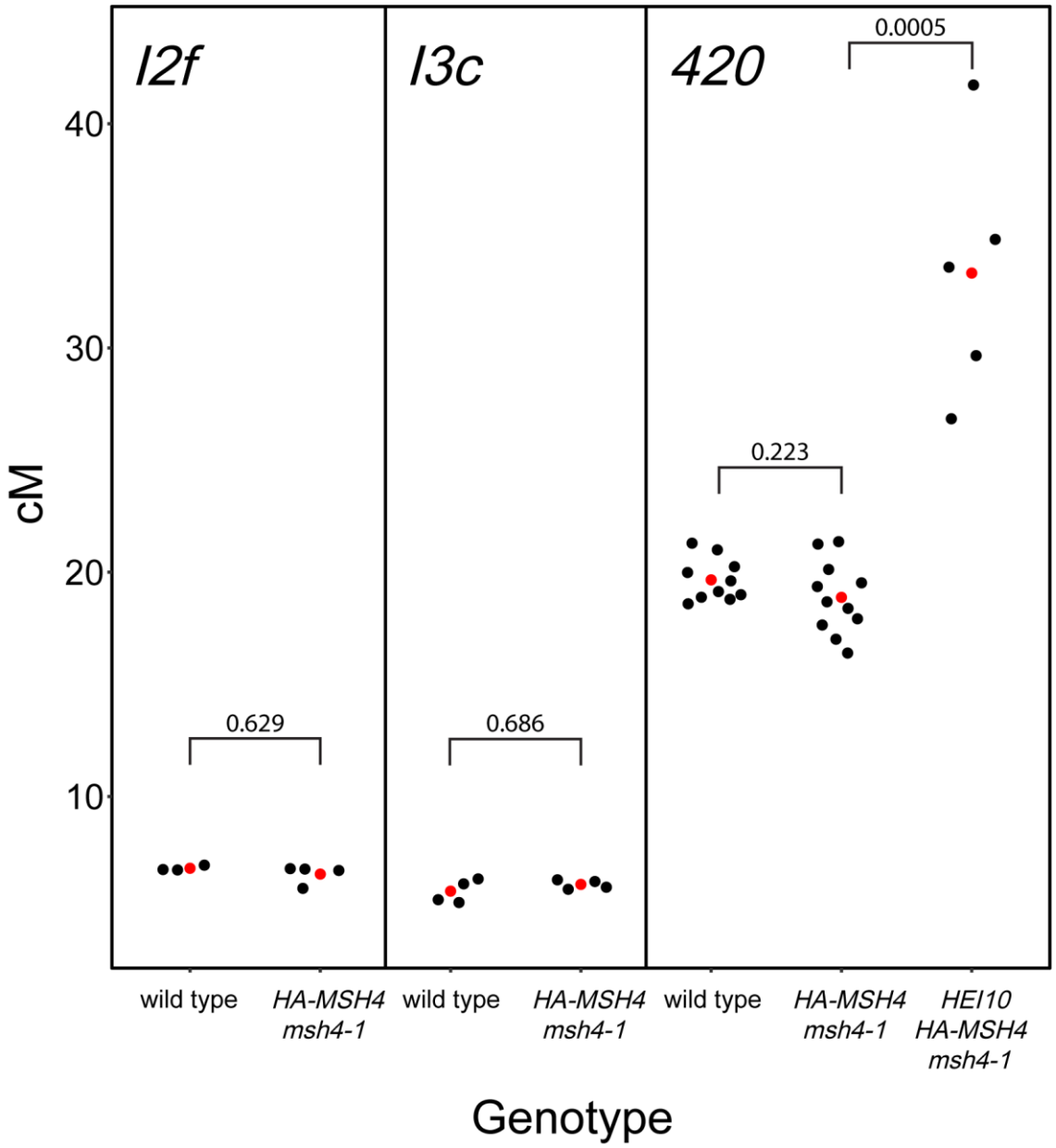


Figure 39. *HA-MSH4 msh4-1* lines have a wild type crossover phenotype.

Genetic distance in centimorgans (cM) at *I2f* and *I3c* was measured in wild type and *HA-MSH4 msh4-1* backgrounds, and at *420* in the wild type, *HA-MSH4 msh4-1, and HEI10 HA-MSH4 msh4-1* backgrounds. Black dots represent replicate measurements and red dots represent the mean of each genotype. Mann-Whitney *U* tests were performed to test for significant differences between genotypes.

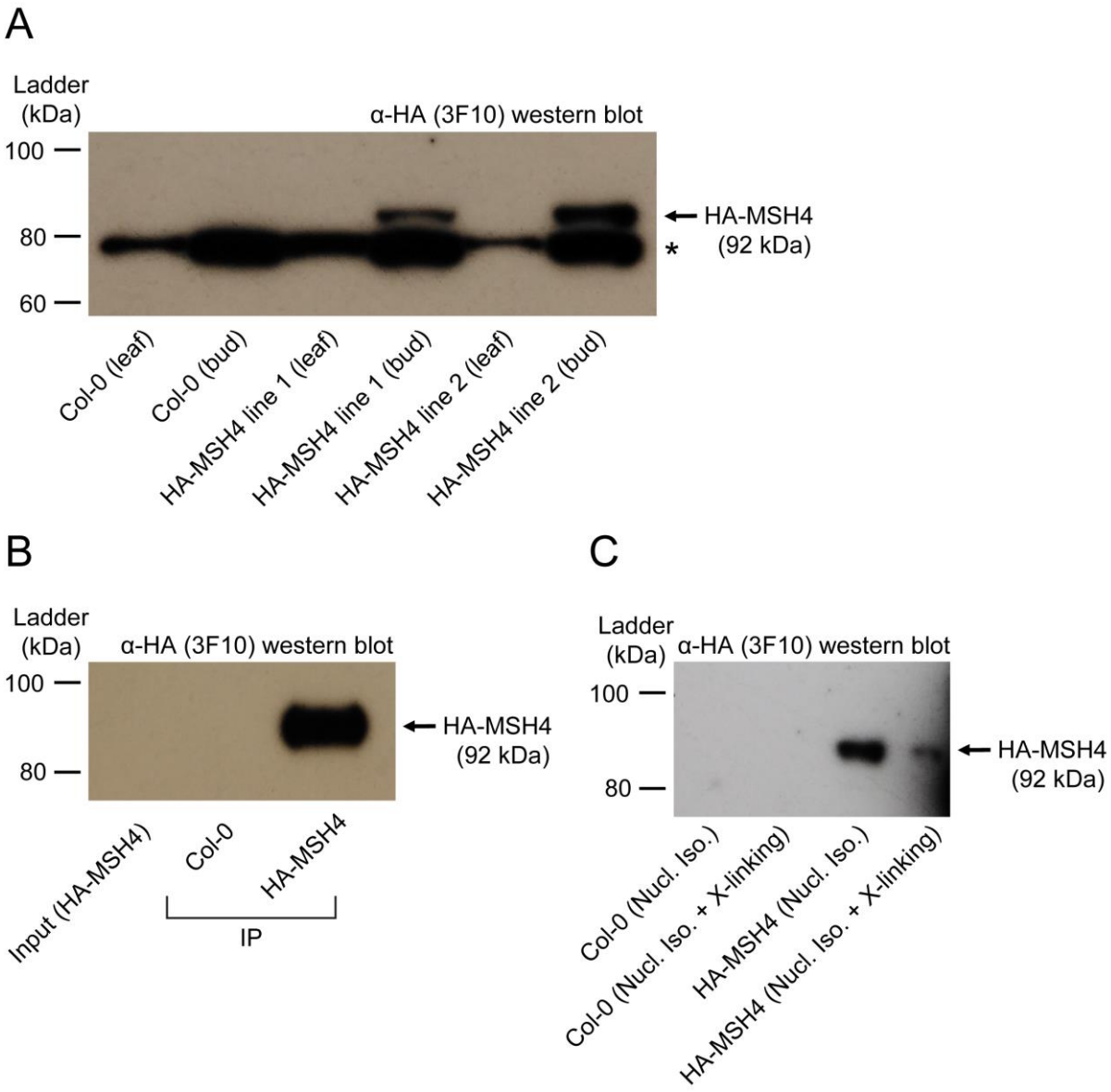


Figure 40. *HA-MSH4* is expressed specifically in bud tissue.

(A) Western blot of protein samples extracted from Col-0 or *HA-MSH4* complementing lines. Protein was extracted from leaves or buds, and 30 µg or 50 µg was loaded per lane, respectively. After transfer, blots were probed with α-HA antibody 3F10 (Roche). A non-specific band, present in all samples, is indicated with an asterisk. The arrow indicates the expected band for HA-MSH4 of ~92 kDa. (B) Western blot after an α-HA IP from *HA-MSH4* or Col-0 bud tissue protein extract. An IP was performed on 9 ml of diluted protein sample, and eluted in a final volume of 50 µl. 10 µl of solution was loaded for the *HA-MSH4* input sample, the negative control (Col-0) IP elution and *HA-MSH4* IP elution, respectively. (C) Western blot of protein samples extracted from nuclear isolates of Col-0 or *HA-MSH4* bud tissue, with or without formaldehyde cross-linking. 10 µl of denatured protein solution was loaded per lane.

5.5 ChIP-qPCR validation of ChIP efficacy.

Before performing a ChIP-seq experiment, I sought to validate the ChIP protocol using antibodies against either H3K4me3 or H3K9me2. Floral bud tissue was ground, crosslinked with formaldehyde and sonicated to enrich for fragments of 150-300 bp. IP was then performed by incubating the soluble chromatin fraction with antibody-bound beads for 16 hours. Chromatin was then washed, crosslinking reversed, and DNA purified. qPCR using either ChIP or input DNA revealed enrichment of H3K4me3 at the 5'-end of a gene known to be actively transcribed and also a crossover hotspot (*RAC1* exon 1) (Serra et al., 2018b) (fig. 41A). In contrast, no enrichment was observed at a heterochromatic transposon locus *Ta2* (Lippman et al., 2003; Yelina et al., 2015). Using an α-H3K9me2 antibody, I observed the opposite relationship: H3K9me2 enrichment was low at the crossover hotspot *3a* (HP3-2, HP4-2), whilst being enriched at the *Ta2* transposon (Yelina et al., 2015) (fig. 41B). These experiments confirmed the efficacy of the ChIP protocol, at least for the enrichment of specific histone modifications.

Next, I sought to investigate the enrichment of MSH4 at several known crossover hotspots (Yelina et al., 2015; Serra et al., 2018b). ChIP was performed by immunoprecipitation of HA-MSH4 from floral bud tissue – as described above, but using a polyclonal ChIP-validated α-HA antibody (ab9110, abcam) – followed by qPCR using previously described primer sets at *3a* (HP3-2, HP4-1, HP4-2), *3b* (HP5-2), and *RAC1* (exon1, exon2) crossover hotspots, or the *Ta2* transposon (Yelina et al., 2015; Serra et al., 2018b). Compared to the input, this analysis revealed that MSH4 was enriched by ~1.5-3-fold at the crossovers hotspots, whilst being less enriched at the *Ta2* transposon (fig. 41C). As these results implied success of the MSH4 ChIP protocol, I proceeded to perform a scaled up ChIP-seq experiment.

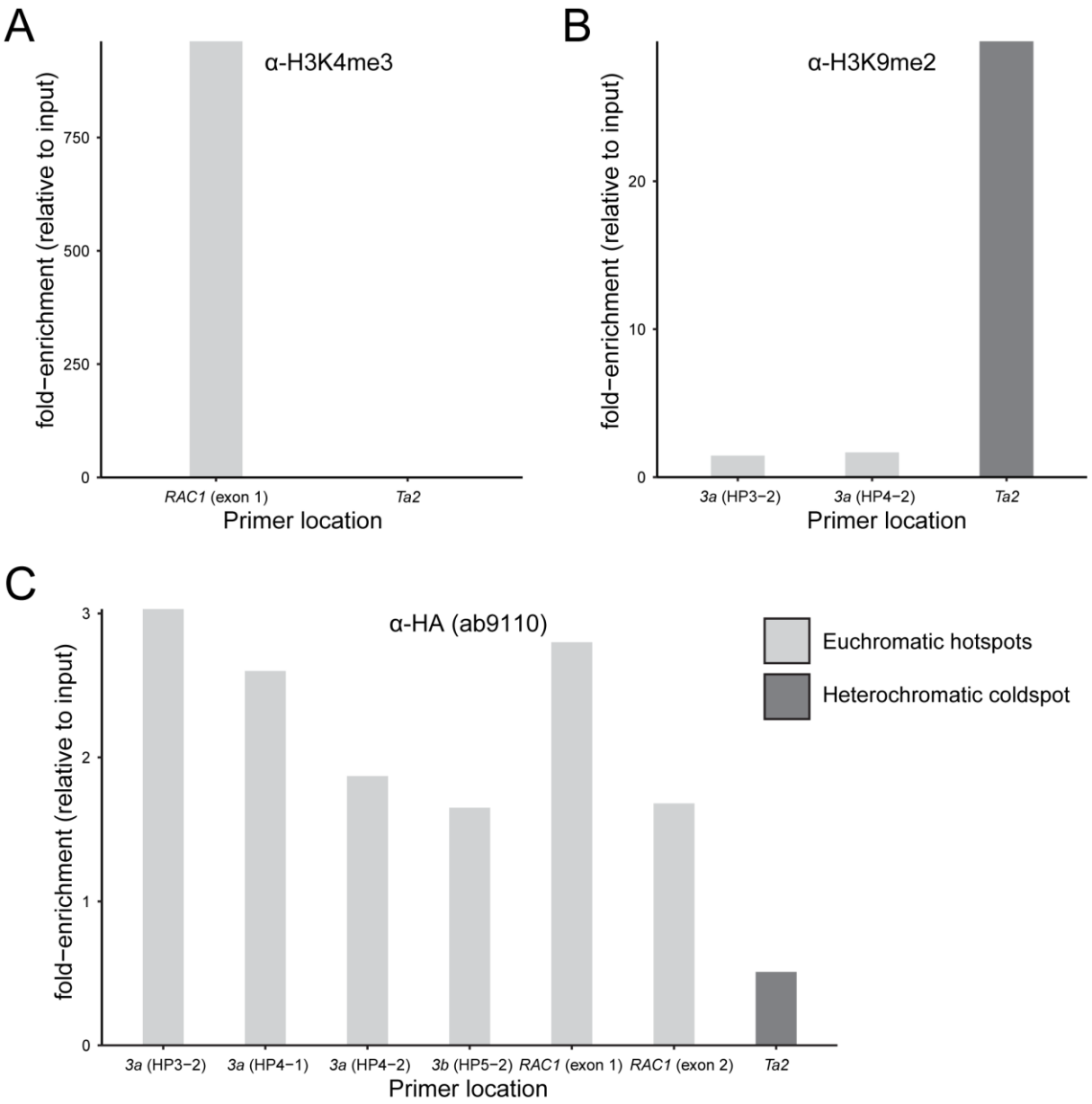


Figure 41. ChIP-qPCR analysis at crossover hotspot and transposon loci.

(A) α-H3K4me3 ChIP-qPCR analysis, using primers sets targeting the *RAC1* (exon 1) or *Ta2* transposon loci. Data are the mean fold-enrichment relative to input DNA from three technical replicates, as calculated from Ct values. (B) As for (A), but with α-H3K4me3 antibody ChIP, and qPCR analyses at the *3a* (HP3-2 and HP4-2) or *Ta2* transposon loci. (C) As for (A), but with α-HA antibody ChIP, and qPCR analyses at the *3a* (HP3-2, HP4-1 and HP4-2), *3b* (HP5-2), *RAC1* (exon 1 and exon 2) or *Ta2* transposon loci.

5.6 MSH4 binding is enriched in the pericentromeres and centromeres.

To map sites of MSH4 binding across the genome, I performed ChIP-seq on 10 g of unopened floral buds collected from the complemented *HA-MSH4* line. These tissues contain all stages of meiosis, and over-represent male compared to female meiocytes (Smyth et al., 1990). Tissue was ground and crosslinked with formaldehyde, before nuclear isolation and lysis. Chromatin was then sonicated to enrich for fragments of 150-300 bp, and HA-MSH4 was immunoprecipitated from the soluble chromatin fraction using α-HA polyclonal antibodies (ab9110, abcam). After washing and reverse-crosslinking, ChIP and input DNA were purified. These fragments were processed into sequencing libraries and 76 bp paired-end sequencing was performed using an Illumina NextSeq instrument.

ChIP and input reads were deduplicated, aligned to the TAIR10 reference genome using Bowtie2 (Langmead & Salzberg, 2012), and filtered for <2 mismatches. Multiple mapping reads were filtered by removing alignments with a MAPQ score lower than 10, and from those remaining only the alignment with the highest MAPQ score was retained. In the cases of identical MAPQ scores, a single alignment was randomly selected. In addition, only read-pairs with both individual reads aligning were retained. This resulted in a final genome coverage of 16.93× and 9.63× for the ChIP and input samples, respectively (appendix 7.18). In parallel, an α-HA ChIP was performed on 10 g of Col-0 unopened floral buds, but in contrast this experiment did not produce enough DNA to make a sequencing library (~7.4 ng total ChIP DNA versus ~12.5 ng of ChIP DNA from the *HA-MSH4* complemented sample). Assuming that these values are reproducible, the reduced quantity of DNA pulled down from the Col-0 ChIP suggested that a significant quantity of the *HA-MSH4* ChIP was specific signal. However, the absence of a α-HA Col-0 negative control sequencing library remains a limitation throughout the subsequent ChIP-seq analysis, and reduces the reliability of the analysis and data interpretation. Ascertaining this negative control remains a priority before publication of this dataset.

These data were then processed with a ChIP-seq analysis pipeline developed by Dr Andrew Tock, for the analysis of REC8-HA ChIP-seq data (Lambing et al., 2019). All data analysis and visualisation was performed by Dr Andy Tock, with active discussion, feedback and intellectual input from myself, Dr Christophe Lambing and Dr Ian Henderson. Before visualising the data, $\log_2$(ChIP/input) enrichment was calculated to control for background DNA and variation in mappability between genomic loci. A Z-score standardisation was applied such that genome-wide mean coverage equals zero, and a value of plus or minus one equates to one standard deviation from the mean. At the chromosome scale, MSH4 enrichment was highly elevated over the pericentromere and centromere, when analysed in sliding 10 kb windows (fig. 42A).

Notably, a similar pattern of enrichment was observed in a parallel ChIP-seq study of the meiotic cohesin subunit REC8 (performed by Dr Christophe Lambing) (Lambing et al., 2019). REC8 is a meiosis specific kleisin subunit of the cohesin complex, a ~35–50 nm ring complex which can topologically embrace sister chromatids after meiotic S-phase (Watanabe et al., 1999; Zickler & Kleckner, 1999; Cai et al., 2003). REC8-cohesin plays a key role in tethering chromatin loops to the meiotic chromosome axis, in addition to promoting ordered polymerization of the axis (Zickler & Kleckner, 1999; Cai et al., 2003; Xu et al., 2005; Lambing et al., 2019). At the chromosome scale, REC8-cohesin was enriched over the centromere and pericentromere and positively correlated with MSH4 (fig. 42A). This positive correlation was observed when comparing the MSH4 ChIP-seq data with REC8 ChIP-seq experiments using either HA- or Myc-tagged REC8 (r_s = 0.93 and 0.77, respectively). Notably, analysing the correlation between MSH4 and REC8 enrichment within the chromosome arms, in 10 kb windows, also revealed a positive correlation (r_s = 0.86). These data indicate that at the chromosome scale and when sampling across meiotic prophase I, MSH4 is predominantly associated with the meiotic axis, as measured by REC8 ChIP-seq.

We then investigated the relationship between MSH4 enrichment and several genomic features, at the chromosome scale. Genome-wide, MSH4 enrichment was negatively correlated with both SPO11-1-oligonucleotide (SPO11-1-oligo) enrichment and a set of 3,320 GBS crossovers, derived from two Col/Ler F_2 populations (r_s = -0.25 and -0.16, respectively) (fig. 42A) (Choi et al., 2018; Serra et al., 2018a; Underwood et al., 2018). Consistently, MSH4 enrichment was also negatively correlated with H3K4me3 ChIP-seq enrichment, a marker of open chromatin, and gene density (r_s = -0.80 and -0.72, respectively) (fig. 42B). These relationships were particularly striking and unexpected, given that genes and open chromatin are relatively enriched for meiotic crossovers (Choi et al., 2013; Serra et al., 2018a), and that MutSγ is known to function as a pro-crossover factor in the ZMM pathway (Higgins et al., 2004, 2008b). Hence, this implies that MSH4 is binding to regions of the genome that do not experience high levels of DSBs or crossovers in wild type.

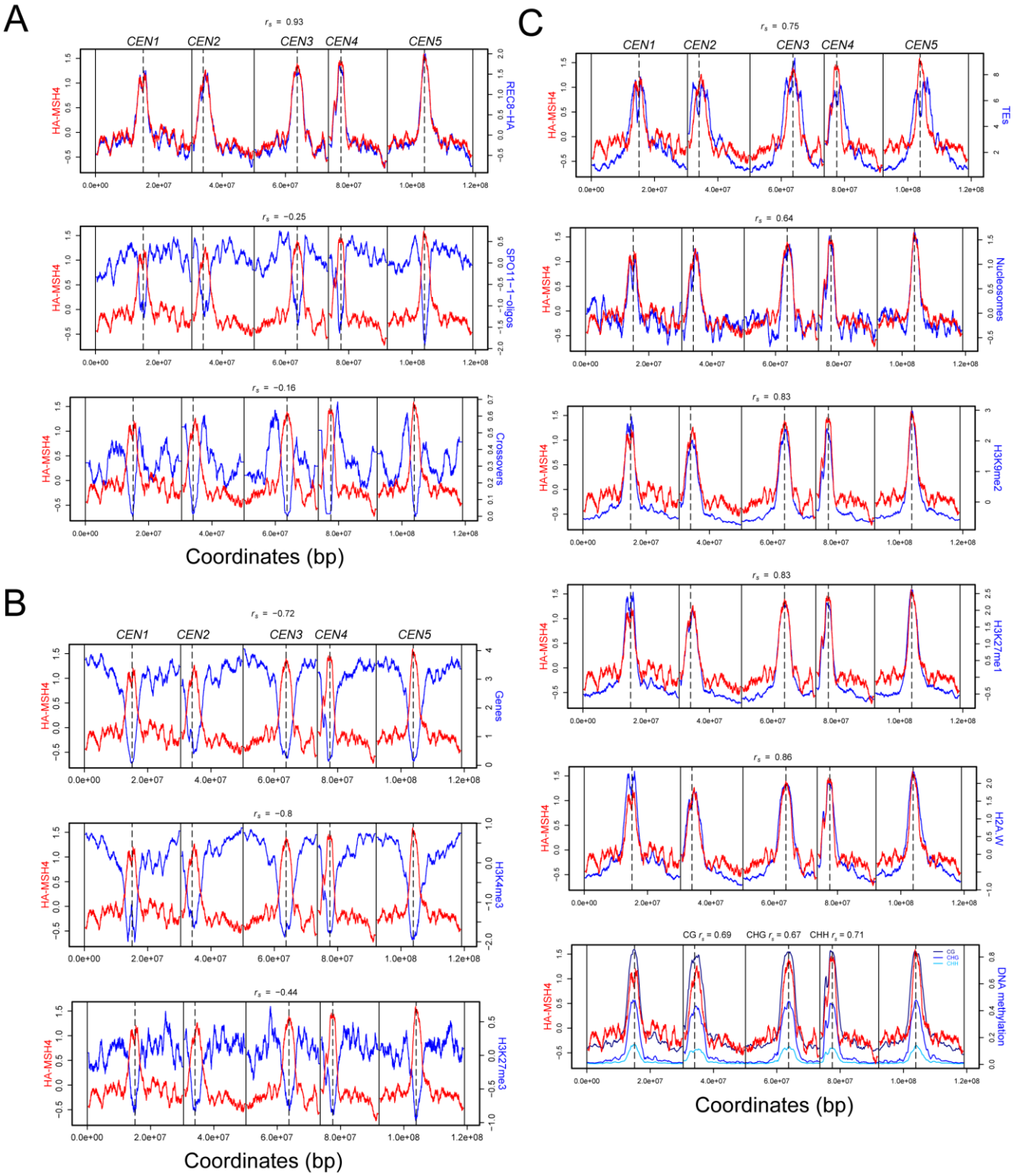

A
B
C
CEN1
CEN2
CEN3
CEN4
CEN5
HA-MSH4
REC8-HA
SPO11-1-oligos
Crossovers
Genes
H3K4me3
H3K27me3
TEs
Nucleosomes
H3K9me2
H3K27me1
H2A.W
DNA methylation
CG
CHG
CHH
Coordinates (bp)

Figure 42. At the chromosome scale, MSH4 is enriched over centromere-proximal chromatin.

(A) Genome wide profiles of log_2(ChIP/input) HA-MSH4 ChIP-seq (red) is contrasted with REC8-HA ChIP-seq (Lambing et al., 2019), SPO11-1-oligos (Choi et al., 2018) and crossover frequency (Serra et al., 2018a; Underwood et al., 2018) (blue), calculated using 10 kb sliding windows. Vertical lines indicate the telomeres and dashed lines indicate centromeres (*CEN1-CEN5*). Spearman's correlation coefficients (r_s) are stated above the plot margins. (B) As for (A), but showing gene density (TAIR10 representative gene models), profiles of H3K4me3 enrichment (Choi et al., 2018), and H3K27me3 enrichment (Lambing et al., 2019). (C) As for (A) but showing transposable element (TE) density (Buisine et al., 2008), profiles of MNase-seq (Choi et al., 2018), and the markers of heterochromatin identity H3K9me2 (Choi et al., 2018), H3K27me1 (Lambing et al., 2019), and H2A.W (Yelagandula et al., 2014) ChIP-seq data sets. % DNA methylation (CG, CHG and CHH) is shown (Stroud et al., 2013).

To assess the contribution of different chromosome regions in driving these genome-wide correlations, a correlation analysis was performed separately on the chromosome arms and the pericentromeres. The pericentromeres were defined as regions adjacent to the centromeres with higher than average DNA cytosine methylation (fig. 42C) (Yelina et al., 2015; Underwood et al., 2018). At the chromosome scale, MSH4 enrichment was positively correlated with nucleosome occupancy genome-wide (r_s = 0.64) (fig. 42A), which was also apparent for the pericentromeres and chromosome arms (r_s = 0.97 and 0.33, respectively). As predicted from the enrichment of MSH4 in proximity to the centromeres, MSH4 was positively correlated with markers of constitutive heterochromatin identity genome-wide: H3K9me2 (r_s = 0.83), H3K27me1 (r_s = 0.83), and H2A.W (r_s = 0.86) (fig. 42). H3K9me2 and H3K27me1 are covalent histone modifications associated with repeat regions and transcriptional silencing (Zhang et al., 2009; Jacob et al., 2009), and H2A.W is a H2A histone variant that associates with H3K9me2 to repress transposon activity and promote chromatin condensation (Yelagandula et al., 2014). H3K9me2 and non-CG DNA methylation forming a self-reinforcing loop (Stroud et al., 2013, 2014), and MSH4 enrichment was positively correlated with CHG and CHH methylation levels genome-wide (r_s = 0.67 and 0.71, respectively), and also CG methylation (r_s = 0.69) (fig. 28C). In contrast, H3K27me3 mediates facultative 'Polycomb' gene repression and occurs largely independent of the other silencing pathways discussed (Jacob et al., 2009). Consistent with its genic enrichment, H3K27me3 was negatively correlated with MSH4 enrichment genome-wide (r_s = -0.44).

Strikingly, the positive association between MSH4 and markers of constitutive heterochromatin was also observed within the chromosome arms: H3K9me2 (r_s = 0.63), H3K27me1 (r_s = 0.62), and H2A.W (r_s = 0.69). Likewise, DNA methylation levels at all three contexts was positively correlated with MSH4 enrichment in the chromosome arms: CG (r_s = 0.47), CHG (r_s = 0.40)

and CHH (r_s = 0.44). However, there was no clear correlation between H3K27me3, a marker of facultative heterochromatin, and MSH4 enrichment in the chromosome arms (r_s = 0.09). As observed for REC8, this indicates that the association between MSH4 and heterochromatic or nucleosome dense regions may not simply be driven by its association with centromere proximal heterochromatin.

5.7 MSH4 is depleted at DNA transposons and enriched at RNA transposons.

Due to the striking correlation with markers of heterochromatin identity, we next assessed the relation between MSH4 enrichment and transposable element (TE) density. At the chromosome scale, MSH4 enrichment was positively correlated with TEs (r_s = 0.75) (fig. 42C). Consequently, we sought to investigate the association between DNA and RNA transposons, which are known to have different chromatin identities and levels of meiotic recombination (Choi et al., 2018). DNA TEs are either autonomous or non-autonomous elements that transpose via DNA intermediates. In contrast, RNA elements amplify themselves and transpose via RNA intermediates. Enrichment of MSH4 was plotted over all DNA or RNA TEs, including 2 kb flanking regions, and compared to a series of genome-wide datasets. As a control comparison, for each dataset enrichment was calculated over a set of random genomic intervals of the same number and width as the TEs. Consistent with the silenced state of TEs, markers of heterochromatin were elevated for both DNA and RNA elements relative to a random comparison, whilst H3K4me3 was depleted (fig. 43A). However, although DNA elements were enriched for H3K9me2, H2A.W and DNA methylation at all three contexts, both MSH4 and REC8 were relatively depleted at these loci and highly correlated with each other (r_s = 0.88) (fig. 43A). Strikingly, despite the increase in heterochromatic markers, nucleosome occupancy was sharply depleted within the DNA TE intervals, and was highly correlated with MSH4 enrichment (r_s = 0.95) (fig. 43A) Consistent with the chromosomal scale anticorrelation between MSH4 and meiotic DSBs (fig. 42A), SPO11-1-oligos are enriched within DNA transposons, as reported (Choi et al., 2018), and anticorrelated with MSH4 enrichment within the TE intervals (r_s = -0.98).

The relationship between MSH4 enrichment and RNA TEs was strikingly different, as previously reported for SPO11-1-oligos and REC8 (Choi et al., 2018; Lambing et al., 2019). At these loci, MSH4 remained highly correlated with REC8 (r_s = 0.85), and negatively correlated with DSBs (r_s = -0.75), but in contrast to DNA TEs MSH4 was enriched over RNA TEs (fig. 43B). Notably, although markers of heterochromatin identity were elevated at these loci, again consistent with their silenced state, nucleosome occupancy was elevated and positively correlated with MSH4 enrichment (r_s = 0.70) (fig. 43B). Together, these comparisons indicate that the distribution of MSH4 at the fine-scale is highly dependent on local genomic

parameters, and, in the context of TEs, suggests that nucleosome density may be a driving factor in determining MSH4 enrichment.

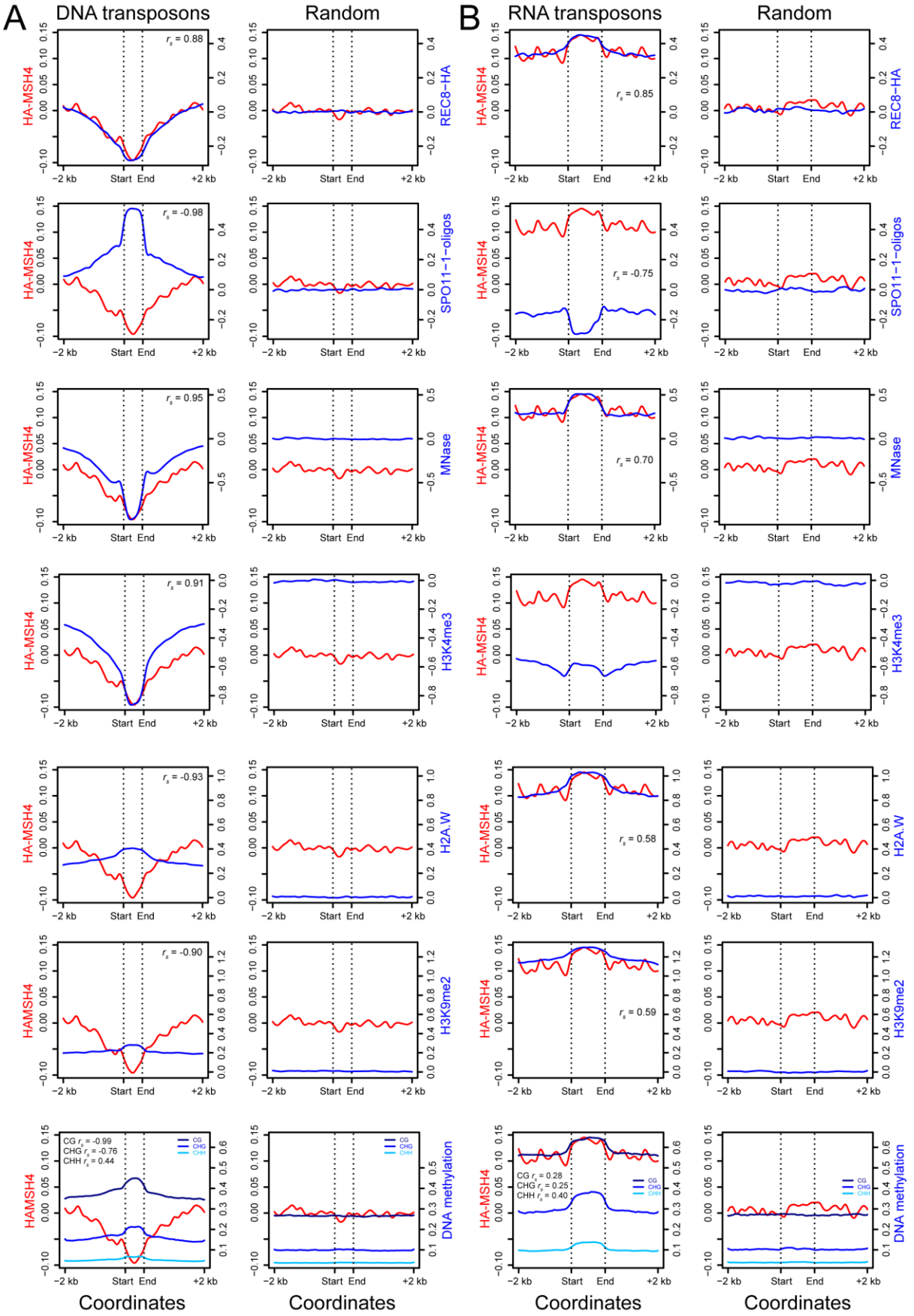

A
DNA transposons
Random
B
RNA transposons
Random
HA-MSH4
REC8-HA
SPO11-1-oligos
MNase
H3K4me3
H2A.W
H3K9me2
DNA methylation
HAMSH4
−2 kb
Start
End
+2 kb
Coordinates
CG
CHG
CHH

Figure 43. MSH4 is depleted at DNA transposons and enriched at RNA transposons.

(A) Log2(ChIP/input) HA-MSH4 enrichment (red) was assessed over DNA transposons, including 2 kb flanking regions, and contrasted with REC8-HA ChIP-seq), SPO11-1-oligos (Choi et al., 2018), MNase-seq, and H3K4me3, H2A.W and H3K9me2 ChIP-seq datasets (Yelagandula et al., 2014; Choi et al., 2018; Lambing et al., 2019) (Buisine et al., 2008). CG, CHG and CHH DNA methylation levels were also contrasted (Stroud et al., 2013). In each case, enrichment was calculated for a set of random regions of the same number and width. Spearman's correlation coefficients (r_s) are stated within the plot margins. (B) As for (A), but for RNA transposons.

5.8 MSH4 is depleted in gene promoters and terminators, but elevated within gene bodies.

We next sought to examine the fine-scale distribution of MSH4 within the genome. Previously, fine-scale associations have been observed between meiotic DSBs and crossovers in relation to plant genes (Giraut et al., 2011; Drouaud et al., 2013; Wijnker et al., 2013; Yelina et al., 2015; Choi et al., 2018; Henderson & Tock, 2018; Serra et al., 2018b). Therefore, we plotted the average enrichment of MSH4 from transcriptional start site (TSS) to transcriptional termination site (TTS) for all TAIR10 representative gene models, including 2 kb upstream or downstream of this window. Because of varying gene lengths, each gene was divided into proportionally scaled windows between start and end coordinates, and 2-kb flanking regions were divided into 20-bp windows. The mean profile, averaged over all genes, was plotted such that the distance between TSS and TSS coordinates along the x-axis represents the mean gene length. As a control comparison, enrichment was calculated for a set of random genomic regions of the same number and width. This analysis procedure was also performed on a set of additional genome-wide datasets for comparison.

MSH4 enrichment appeared relatively depleted in gene promoters immediately upstream of the TSS, but then increased throughout the gene body, reaching a peak of enrichment just upstream of the TTS (fig. 44). This result was surprising, given our understanding of MSH4 as a pro-crossover protein and both experimental and historical measurements of meiotic recombination indicating that DSBs and crossovers are enriched towards gene promoters and terminators, and that crossover hot spots are associated with active chromatin modifications located at gene 5′-ends, including H2A.Z and H3K4me3 (Choi et al., 2013; Drouaud et al., 2013; Wijnker et al., 2013; Yelina et al., 2015; Choi et al., 2016; Henderson & Tock, 2018; Serra et al., 2018b). Strikingly, the TSS-TTS distribution of MSH4 was highly similar to that of the meiotic cohesin REC8 (r_s = 0.98), suggesting that these distributions may be indicative of a meaningful relationship between gene structure and meiotic recombination proteins, and suggest that MSH4 is associating with the chromosome axis during meiotic recombination. As

predicted from the positive associated with nucleosomes, the distribution of MSH4 within genes differed greatly to the pattern of SPO11-1-oligos, and showed almost no correlation between the TSS-TTS (r_s = -0.07). SPO11-1-oligos are highly enriched 5′ and 3′ of the nucleosome depleted TSS and TTS, respectively, and depleted within the gene body (fig. 44) (Choi et al., 2018).

The MSH4 distribution was then compared to a series of chromatin features. Strikingly, the distribution of nucleosomes across genes, as measured by MNase-seq, correlated strongly with MSH4 (r_s = 0.90), except immediately downstream of the TSS where nucleosome density was higher, peaking at the +1 nucleosome (fig. 44). However, nucleosome occupancy showed a similar trend of enrichment when moving towards the 3′ of the gene body (fig. 44). Consistent with the chromosome scale correlations, MSH4 was anticorrelated with H3K4me3 within the TSS-TTS window (r_s = -0.69), which in contrast to MSH4 showed the highest level of enrichment immediately downstream of the TSS (fig. 44). As expected, the heterochromatic marks H3K9me2, H3K27me1 and H3K27me3 were depleted between gene TSS-TTS (fig. 44). This is expected given the depletion of genes in the heterochromatin dense centromere proximal regions, where these marks are highly enriched (fig. 42). However, the distributions of all three chromatin modifications were positively correlated with MSH4 enrichment within the TSS-TTS window, despite their low level (r_s = 0.84, 0.54 and 0.44, respectively), in addition to the histone variant H2A.W (r_s = 0.98) (fig. 44). In summary, at the level of genes we observe enrichment of MSH4 binding within gene bodies, in a pattern that strikingly correlates with nucleosomes and REC8-cohesin.

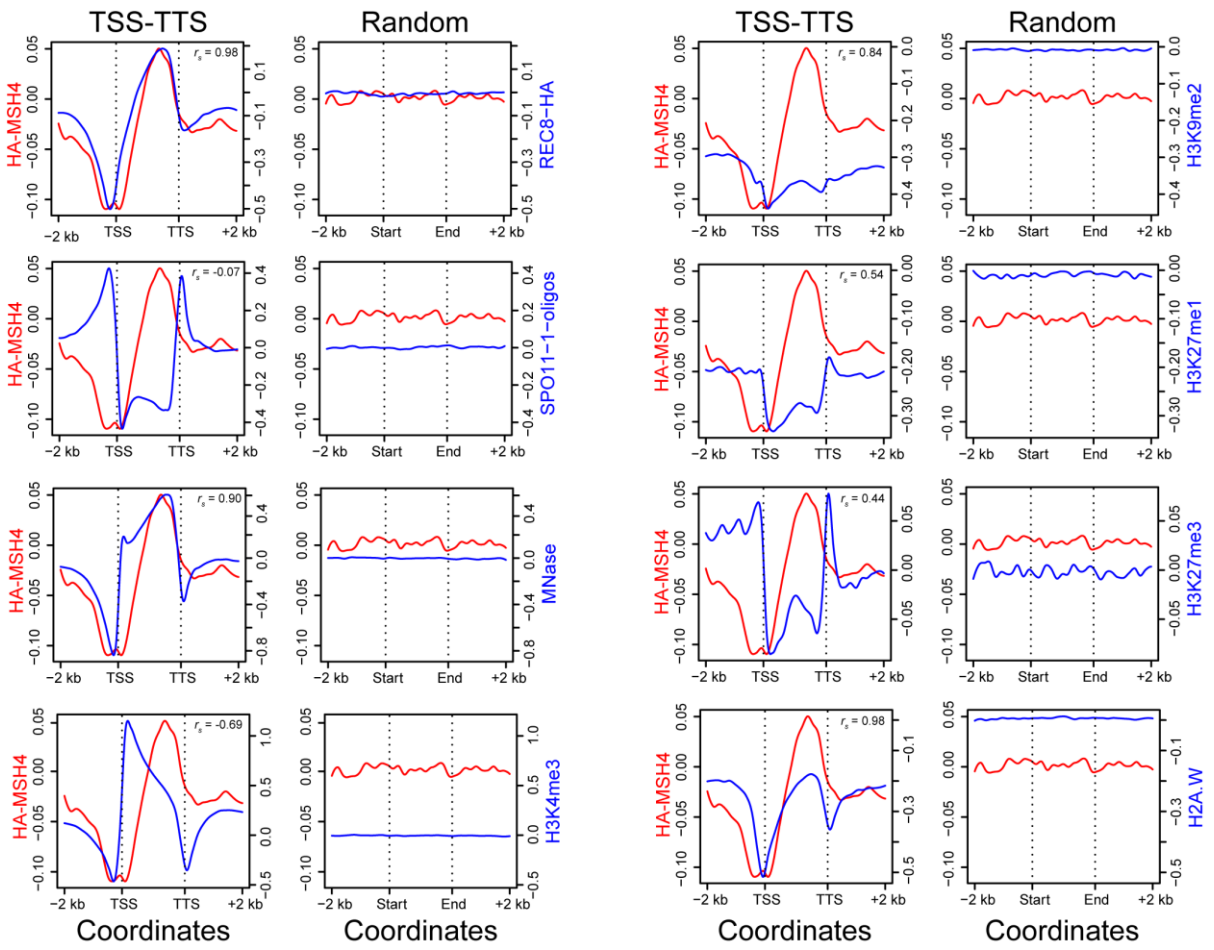


Figure 44. Analysis of MSH4 ChIP-seq enrichment within TSS-TTS windows, and a comparison with additional genomic features.

Plots of HA-MSH4 $\log_2$(ChIP/input) enrichment (red) within transcription start site (TSS) and transcription termination site (TTS) windows for all genes, including 2 kb flanking regions, compared with REC8-HA ChIP-seq, SPO11-1-oligos, MNase-seq, and H3K4me3 ChIP-seq (blue) (left column) and H3K9me2, H3K27me1, H3K27me3, and H2A.W ChIP-seq datasets (blue) (right column) (Yelagandula et al., 2014; Choi et al., 2018; Lambing et al., 2019). In each case, enrichment was calculated for a set of random regions of the same number and width. Spearman's correlation coefficients (r_s) are stated within the plot margins.

5.9 MSH4 is enriched within exons and depleted within introns.

We next assessed the relationship between MSH4 enrichment and exon versus intron positions. For this analysis, comparison was focused on MSH4, REC8 and nucleosomes. Average coverage was calculated across exonic or intronic regions, oriented 5′-3′ relative to transcription and normalised to control for differences in exon/intron length. As a control, a set of random genomic intervals of the same widths were compared. These distributions were plotted including plus and minus 1 kb of flanking sequence. This revealed that MSH4 binding was relatively enriched across exons, and depleted within introns, compared to the adjacent sequences (fig. 45). A near identical pattern was observed for REC8 coverage across exons and introns (r_s = 0.74 and 0.98, respectively) (fig. 45). Nucleosome occupancy was likewise enriched within exons and was positively correlated with MSH4 enrichment (r_s = 0.74), but was more distinctly increased relative to adjacent sequences (fig. 45A). Likewise, nucleosomes were more depleted within exons, relative to both MSH4 and REC8, and in fact showed a slight negative correlation with MSH4 (r_s = -0.14) (fig. 45B). It should be noted that MSH4 enrichment was relatively reduced upstream of the exon window, likely caused by the depletion of MSH4 towards gene promoters, and relatively increased downstream of the window due to the enrichment of MSH4 within gene bodies towards the TTS. This phenomenon was also observed for REC8 and MNase data, and for the intron analysis. Therefore, this fine-scale analysis, in relation to gene organisation, again indicated positive associations between MSH4, REC8-cohesin, and nucleosomes.

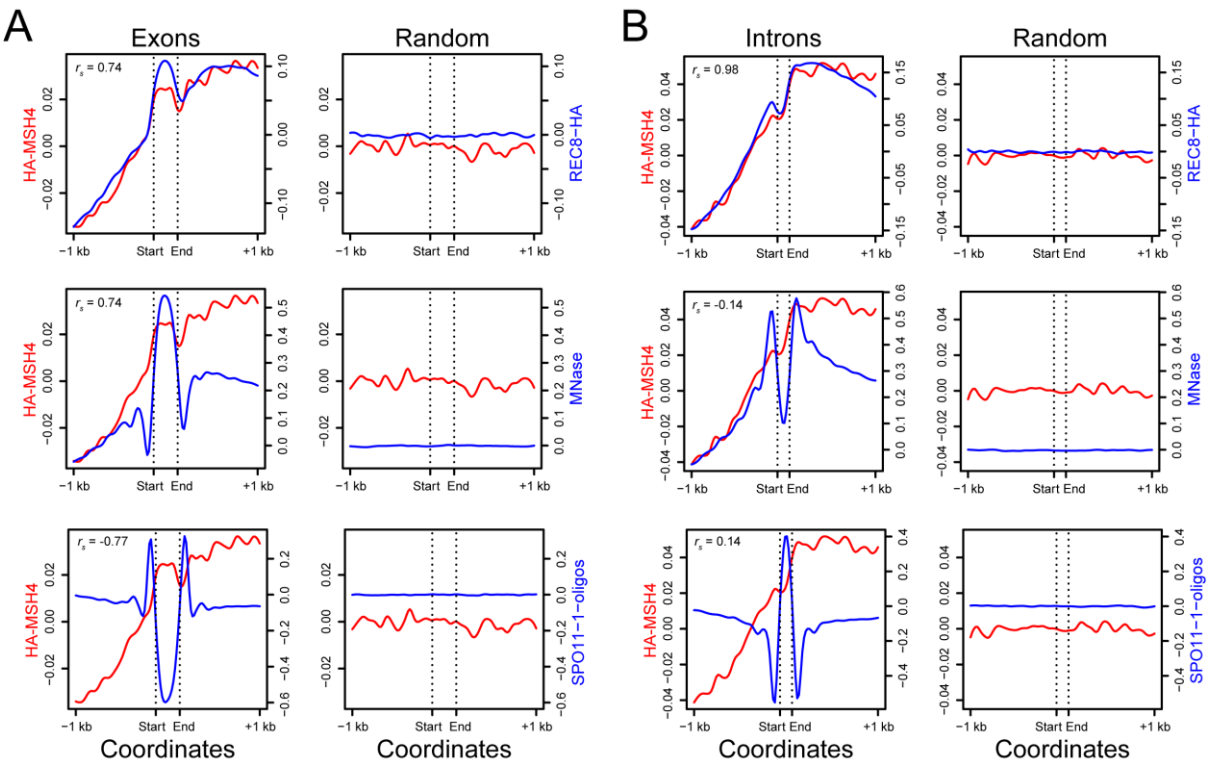


Figure 45. MSH4 is enriched within exons and depleted within introns.

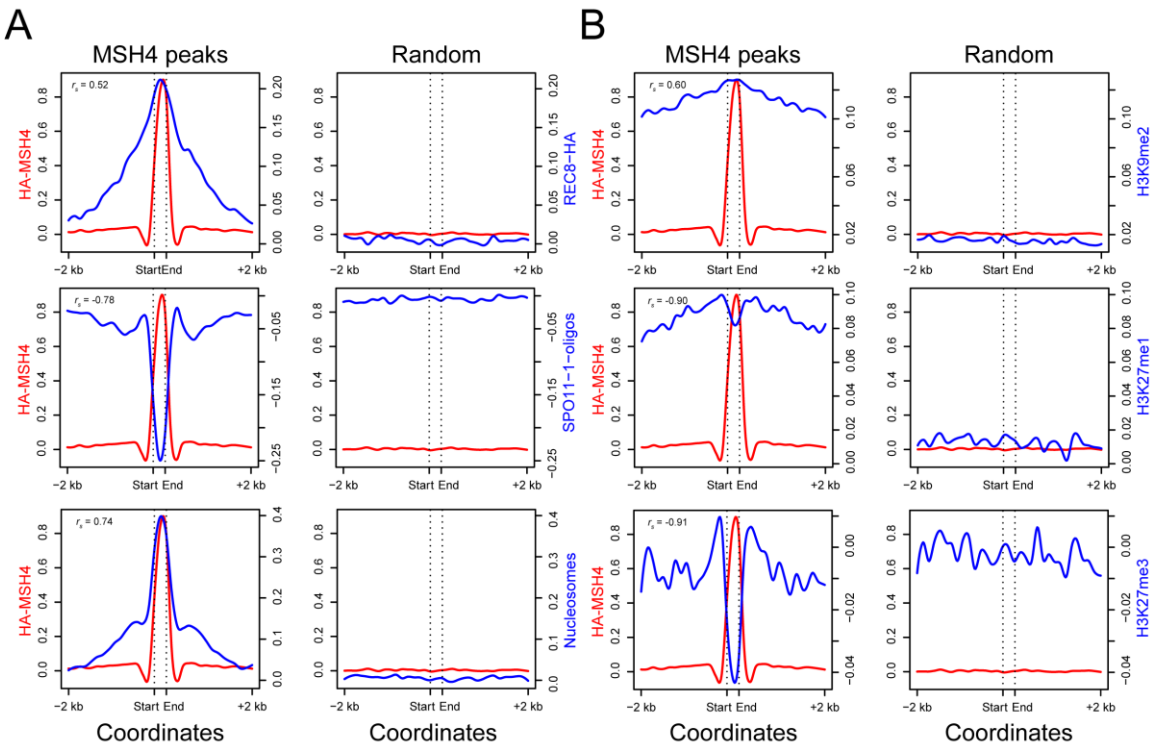
(A) Plots of HA-MSH4 $\log_2$(ChIP/input) enrichment (red) within exons for all genes, including 1 kb flanking regions, compared with REC8-HA ChIP-seq, SPO11-1-oligos and MNase-seq (blue) (Choi et al., 2018; Lambing et al., 2019). In each case, enrichment was calculated for a set of random regions of the same number and width. Spearman's correlation coefficients (r_s) are stated within the plot margins. (B) As for (A), but for introns windows.

5.10 Peaks of MSH4 enrichment are associated with REC8-cohesin and depleted for SPO11-1-oligos.

To further investigate MSH4 patterns at the fine-scale, ChIP-seq peaks were identified using the peak-calling 'ranger tool' within the PeakRanger suite, which first identifies coverage-enriched regions and then identifies local summits of high read density (Feng et al., 2011). In order to avoid peaks being called due to variability in mapping between genomic loci or artefacts from the alignment process, peaks of MSH4 enrichment were defined relative to the input DNA enrichment profile. As MSH4 is broadly associated with chromatin across the genome, stringent significance thresholds ($P \leq 0.001$ and FDR ≤ 0.01) were applied for the detection of peaks. This analysis produced a total of 40,556 MSH4 peaks genome-wide, with an average width of 293 bp. As a control comparison, 40,556 random genomic loci of matched number and width distribution were generated. Plotting the coverage of REC8-HA ChIP-seq, MNase-seq and SPO11-1-oligos across the MSH4 peak regions showed an enrichment of REC8-cohesin and nucleosomes (r_s = 0.52 and 0.74, respectively), and a corresponding depletion of SPO11-1-oligos (r_s = -0.78) (fig.46A), as expected from previous studies (Choi et al., 2018; Lambing et al., 2019). Consistent with the chromosome scale positive correlation between MSH4 and markers of heterochromatin identity, MSH4 peaks were elevated in H3K9me2 and H3K27me1 levels relative to the genome average, whilst being depleted in H3K27me3, a histone modification associated with gene silencing and hence depleted from centromere proximal chromatin (fig. 46B). However, within the peak window, only H3K9me2 was positively correlated with MSH4 enrichment (r_s = 0.60), whereas K3K27me1 and H3K27me3 were negatively correlated (r_s = -0.90 and -0.91, respectively).

In order to further dissect this negative relationship with SPO11-1-oligos, MSH4 peak overlap was assessed across TE superfamilies, previously shown to vary in meiotic DSB levels (Choi et al., 2018). For this analysis, permutation tests were performed using the regioneR package, to assess the significance of the observed overlap between MSH4 peaks or SPO11-1-oligo hotspots and a given TE superfamily (Gel et al. 2016; Choi et al., 2018; Lambing et al., 2019). These values were compared to the average overlap calculated from 10,000 sets of random

genomic loci of matched width distributions. Consistent with the MSH4 peak profiles (fig. 46A), in all cases where fewer-than-expected SPO11-1-oligo hotspots overlapped with a given TE superfamily, MSH4 peaks overlapped that superfamily significantly more-than-expected. The inverse relationship was also observed, with the exception of the hAT superfamily, where fewer-than-expected MSH4 peaks overlapped this superfamily despite SPO11-1-oligos showing no significant enrichment or depletion (fig. 46C). Together with the TSS-TTS analyses, these MSH4 peak analyses indicate that at the fine-scale MSH4 enrichment anticorrelates with meiotic DSBs across multiple genomic contexts.



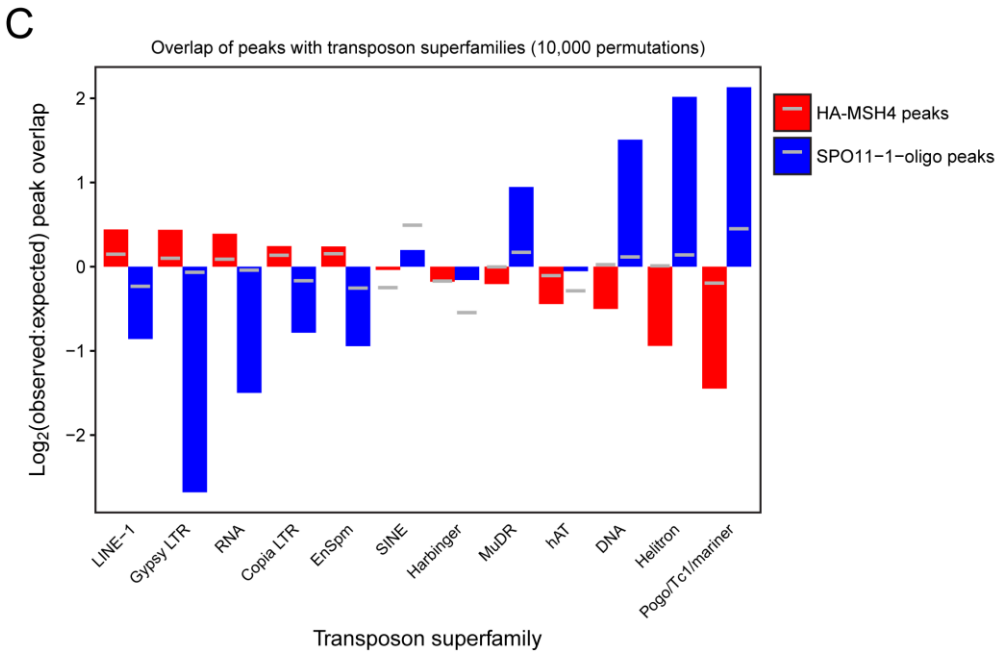


Figure 46. MSH4 peaks are associated with meiotic cohesin, nucleosomes and markers of constitutive heterochromatin, and anticorrelate with meiotic DSBs at transposon superfamily classes.

(A) Profiles of log_2(ChIP/input) HA-MSH4 (red) averaged across 40,556 MSH4 peaks, contrasted with REC8-HA ChIP-seq (Lambing et al., 2019), SPO11-1-oligos (Choi et al., 2018), and MNase-seq (nucleosomes) (Choi et al., 2018) (blue). In each case, random positions of the same number and width were plotted. Plots include 2 kb flanking regions. (B) As for (A), but contrasting HA-MSH4 (red) with H3K9me2 (Choi et al., 2018), H3K27me1 and H3k27me3 (Lambing et al., 2019) ChIP-seq datasets (blue). (C) Bar graphs showing the ratio of observed:expected peak overlap from permutation tests assessing the overlap of HA-MSH4 and SPO11-1-oligo peaks with transposon superfamily classes, based on 10,000 permutations. Horizontal grey lines indicate a significance threshold of $\alpha = 0.05$.

5.11 Ranking genes by MSH4 enrichment reveals a relationship to transcriptional activity.

To further investigate the relationships identified through analysis of averaged MSH4 enrichment across genes, we next sought to subdivide the ChIP-seq data based on enrichment level within TSS-TTS and test for correlations with other features. Genes were divided into six groups (hexiles), with genes ranked according to log_2(MSH4 ChIP/input) levels between their TSS and TTS. For the highest MSH4 hexile, coverage was elevated throughout the TSS-TTS region, and lacked the 5′-3′ polarisation seen for the averaged TSS-TTS profile (fig. 47A). In contrast, there was a noticeable enrichment of MSH4 towards the TTS for all other hexiles (fig. 47A). For the lowest MSH4 hexile, MSH4 was significantly depleted throughout the TSS-TTS, relative to the adjacent DNA sequences (fig. 47A). Ranking all genes relative to MSH4 coverage revealed that as total MSH4 level is reduced, the protein's distribution is increasingly shifted towards the 3′ of the gene body (fig. 47C).

We compared the distribution of several other datasets across TSS-TTS, divided into hexiles ranked by MSH4 coverage. This analysis enables the identification of genomic parameters associating with relative MSH4 enrichment or depletion. Firstly, dividing REC8 enrichment into hexiles based on MSH4 coverage showed that MSH4 and REC8 were positively associated, consistent with the previous chromosome scale and fine-scale analyses (fig. 47A). Strikingly, there was no relationship between MSH4 binding and SPO11-1-oligo density, with all six hexiles having equivalent distributions. This suggests that the meiotic recombination protein MSH4 and meiotic axis protein REC8 are enriched within the same genes, and that the localisation of MSH4 is not correlated with meiotic DSB levels in gene promoters and terminators, and *vice versa*.

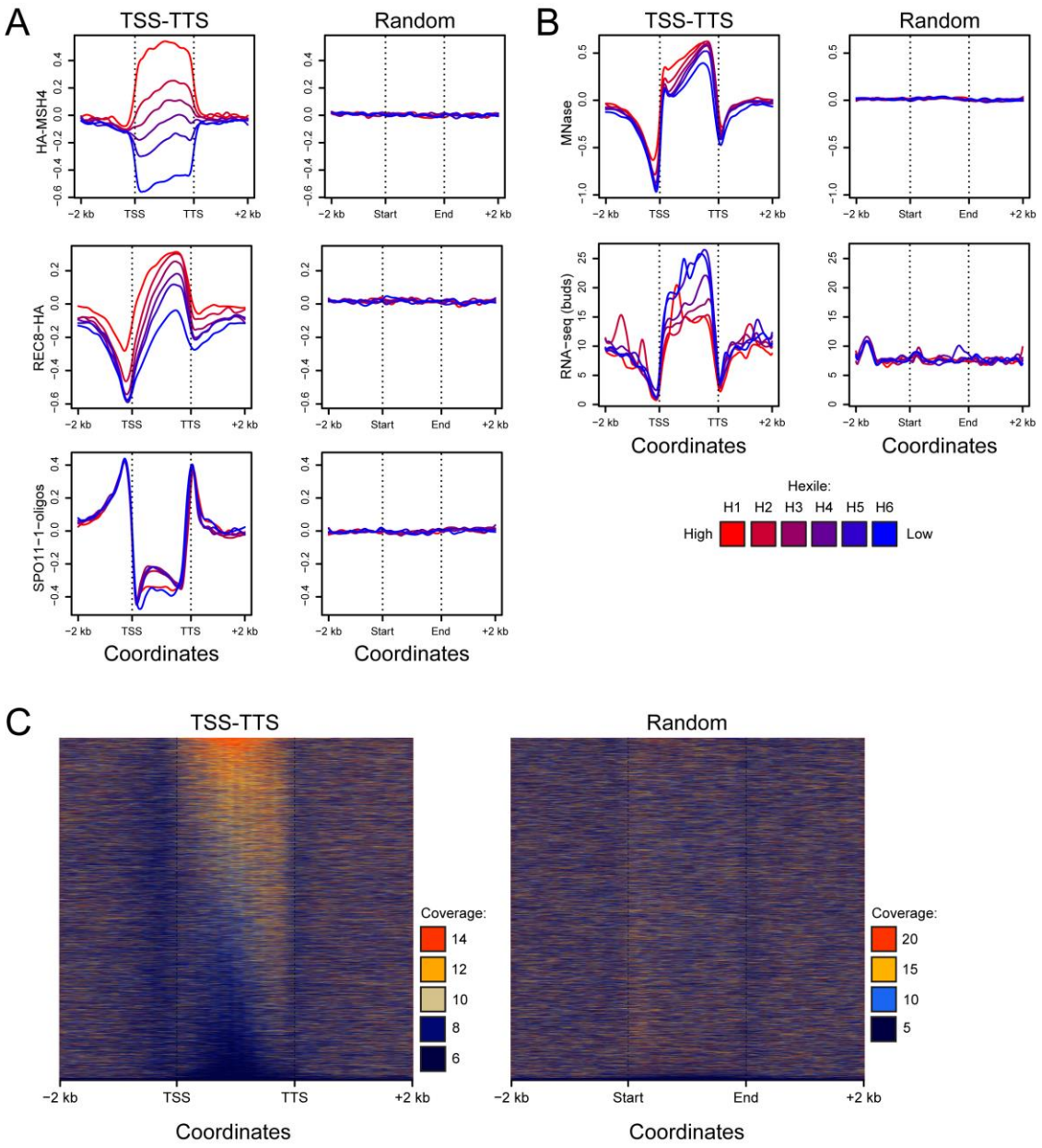


Figure 47. MSH4 enrichment within genes is positively correlated with meiotic cohesin and nucleosomes, and negatively correlated with transcription.

(A) Genes were divided into hexiles, ranked by HA-MSH4 $\log_2$(ChIP/input) enrichment between the TSS-TTS (red = highest, blue = lowest). HA-MSH4 $\log_2$(ChIP/input) enrichment, REC8-HA ChIP-seq and SPO11-1-oligos for each hexile are plotted across the TSS-TTS window, including 2 kb flanking windows (Choi et al., 2018; Lambing et al., 2019). In each case, enrichment was calculated for a set of random regions of the same number and width. (B) As for (A), but plotting MNase-seq and RNA-seq from floral buds for each HA-MSH4 hexile (Choi et al., 2018). (C) All genes were ranked by HA-MSH4 $\log_2$(ChIP/input) enrichment between the TSS-TTS window, from top to bottom, and HA-MSH4 enrichment was plotted from TSS-TTS including 2 kb flanking regions. Enrichment was calculated for a set of random regions of the same number and width.

Next, we compared MSH4 enrichment to RNA-seq data from Arabidopsis floral bud tissues (Choi et al., 2018). Notably, dividing RNA-seq levels into hexiles ranked by MSH4 enrichment within genes revealed a negative relationship between MSH4 and transcription. In particular, the two lowest hexiles of MSH4 enrichment had a pronounced increase in RNA-seq levels relative to the four highest hexiles (fig. 47B). To further investigate the relationship between MSH4 enrichment and transcription, we plotted markers of euchromatin or heterochromatin across TSS-TTS, likewise divided into hexiles based on MSH4 coverage.

Consistent with the negative relationship between RNA-seq levels and MSH4, the MSH4 hexiles displayed a negative relationship with the marker of active transcription H3K4me3 (fig. 48A). In contrast, the highest enriched MSH4 hexile was noticeably elevated in the heterochromatic marks H3K9me2, H3K27me1, H3K27me3 and H2A.W (fig. 48A). This observation indicates that MSH4 is enriched at both constitutively (H3K9me2, H3K27me1 and H2A.W) and facultatively (H3K27me3) silenced genes, which share the feature of suppressed RNA polymerase II transcription. In addition, there was a particularly strong association between the highest MSH4 hexile and CHG and CHH methylation, known to be markers of constitutive transcriptional repression (fig. 48B). Interestingly, a less clear relationship with CG methylation was observed, which may be due to this mark being enriched within the gene bodies of many actively transcribed genes (i.e. gene body methylation), as well as high levels of CG methylation being associated with gene and TE silencing when combined with CHG and CHH context methylation (Bewick & Schmitz, 2017; Zilberman, 2017). In summary, MSH4 hexile analysis suggests that those genes with highest levels of MSH4 binding are transcriptionally repressed, and that repression is associated with elevated CHG and CHH methylation, and histone modifications and histone variants associated with transcriptional repression. Furthermore, this relationship is consistent with the chromosome scale positive correlation between MSH4 enrichment and TE density (fig. 42).

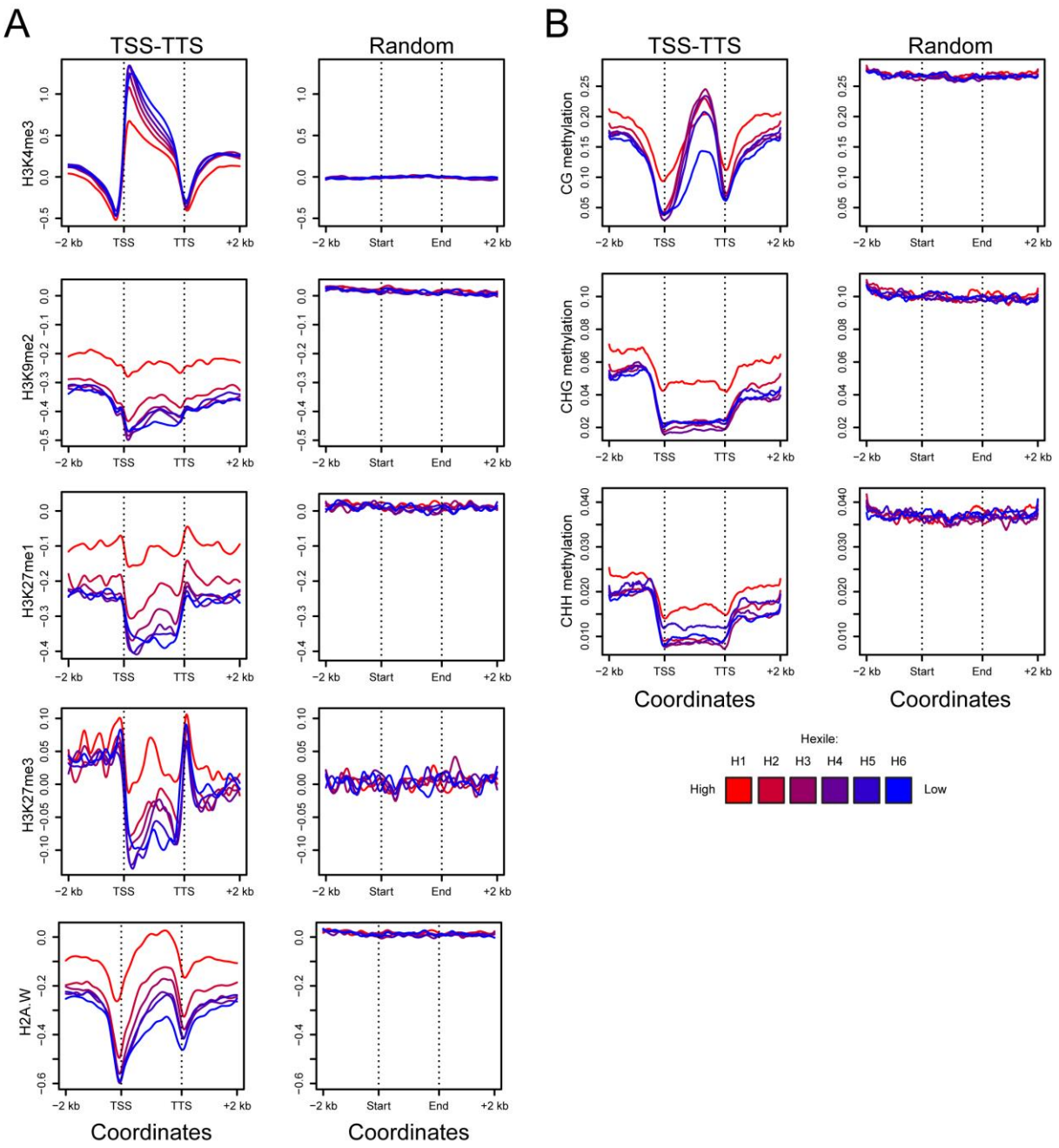


Figure 48. MSH4 enrichment within genes is positively correlated with markers of heterochromatin and negatively correlated with markers of euchromatin.

(A) Genes were divided into hexiles, ranked by HA-MSH4 $\log_2$(ChIP/input) enrichment between the TSS-TTS. The enrichment of H3K4me3, H3K9me2, H3K27me1, H3K27me3 and H2A.W is plotted from TSS-TTS for each hexile (red = highest, blue = lowest), including 2 kb flanking regions (Yelagandula et al., 2014; Choi et al., 2018; Lambing et al., 2019). In each case, enrichment was calculated for a set of random regions of the same number and width. (B) As for (A), but showing DNA methylation level at CG, CHG and CHH contexts within each HA-MSH4 hexile (Stroud et al., 2013).

5.12 MSH4 enrichment is depleted at sites of meiotic crossover.

To investigate the relationship between crossover positions and MSH4 enrichment, MSH4 was assessed within 3,320 GBS crossover intervals, derived from Col/Ler wild type F_2 populations (Serra et al., 2018a; Underwood et al., 2018). Crossover intervals were defined by the 5' and 3' SNPs used to define the change in genotype associated with crossover junctions within the F_2 individuals. As the distance between adjacent SNPs varies, MSH4 enrichment was normalised by interval width and plotted across the mean interval width, including 5 kb flanking regions. As a control, MSH4 enrichment was calculated within a set of random genomic intervals of matched widths. This revealed a clear reduction in MSH4 enrichment in proximity to sites of meiotic crossover (fig. 49). Consistent with the relationships observed in other genomic contexts, REC8 was similarly reduced around crossover intervals, whilst SPO11-1-oligos were elevated (fig. 49). These relationships further show a surprising local negative correlation between MSH4 enrichment and crossover sites, despite its pro-crossover role during meiosis.

5.13 MSH4 anticorrelates with well-positioned nucleosomes.

To further assess the relationship between MSH4, REC8-cohesin and nucleosomes at the fine-scale, we plotted ChIP-seq enrichment over ‘well-positioned’ nucleosomes (WPNs). WPNs are a subset of nucleosomes that show a relatively invariant position, in relation to the underlying DNA sequence (Zhang et al., 2015). In Arabidopsis, a previous study found that ~12% of nucleosomes in Arabidopsis fitted this criteria, and were often associated with gene TSSs (Zhang et al., 2015). Here, we identified WPNs with the R package nucleR, by analysing the ratio of MNase-seq coverage to genomic DNA sequencing coverage at each genomic coordinate, after normalising by library size (Flores & Orozco, 2011). Paired-end 100-bp MNase-seq reads were trimmed to the central 40 bp, which improves nucleosome dyad detection, sharpens the MNase-seq peaks, and reduces the effects of MNase bias towards certain dinucleotide sequences (Deniz et al., 2011; Flores & Orozco, 2011). Paired-end reads from a Col-0 gDNA library were processed in the same way, and provided a control for identifying nucleosome dyad peaks (Flores & Orozco, 2011). This produces $\log_2$(MNase-seq/gDNA-seq) coverage ratios at each genomic coordinate. These values were fast Fourier transformed to remove noise, by retaining 2% of the components of the original profiles, and peaks in the signal were detected which were centred on a local maximum and had a width of 140 bp (Choi et al., 2018). Peaks were then assigned height scores with nucleR, and filtered to retain only the highest confidence peaks (height > 0.99), and overlapping peaks were merged. This resulted in a final subset of 57,734 WPNs, with an average width of 145 bp.

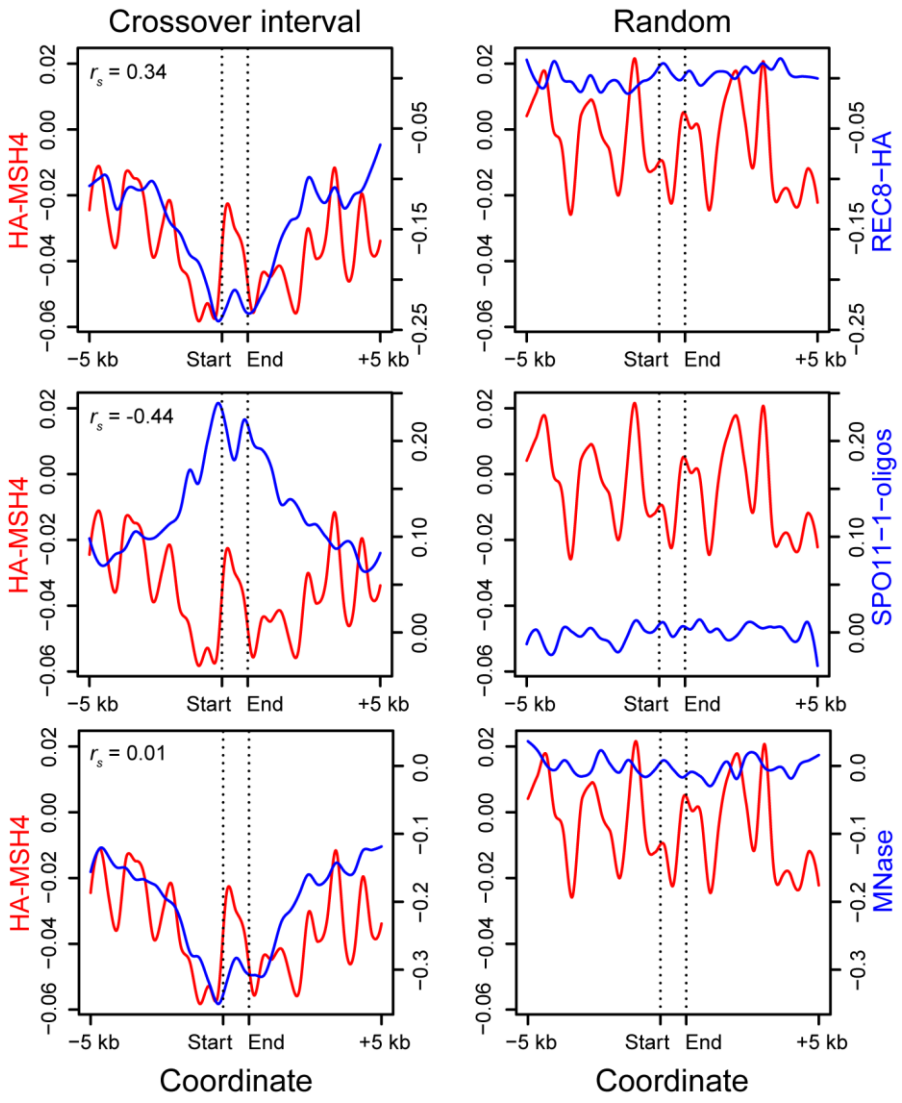


Figure 49. HA-MSH4 and REC8-HA are depleted at sites of meiotic crossover, whilst DSBs are elevated.

Plots of HA-MSH4 $\log_2$(ChIP/input) enrichment (red) within 3,320 GBS crossover intervals, defined by 5' and 3' SNPs demarcating the crossover position, including 5 kb flanking regions (Serra et al., 2018a; Underwood et al., 2018). HA-MSH4 enrichment is compared with REC8-HA ChIP-seq, SPO11-1-oligos and MNase-seq (blue) (Choi et al., 2018; Lambing et al., 2019). Spearman's correlation coefficients (r_s) are stated within the plot margins.

As expected, the sequence composition at WPN sites was highly biased towards GC-rich sequences (fig. 50B), which are known to associate with stable nucleosome binding (Lorch et al., 2015). MNase-seq enrichment was then plotted across WPN sites, centred on the midpoint of their start and end coordinates, including 500 bp of flanking sequence. MSH4 coverage at these positions was overlaid, which revealed a striking distribution (fig. 50A). Consistent with previous analyses, MSH4 enrichment was elevated over the WPN regions, relative to the random control comparison (fig. 50A). However, closer inspection revealed that MSH4 enrichment appeared to oscillate across the window, with a periodicity that locally anticorrelated with peaks of nucleosome occupancy, both at the WPN themselves and in the flanking regions (fig. 50A). Hence, within the WPN window MSH4 enrichment was negatively correlated with MNase-seq enrichment (r_s = -0.96). Strikingly, this periodic relationship was not observed for REC8 ChIP-seq enrichment, which was highly elevated over the centre of the well-positioned nucleosome interval and reduced on either side (fig. 50A), and REC8 enrichment was also positively correlated with MNase-seq within the WPN window (r_s = 0.96). Hence, this analysis potentially reveals that the fine-scale distribution of MSH4 may be biased to inter-nucleosomal linker sequences to a greater degree than the REC8-cohesin complex.

Consistent with a positive association between markers of heterochromatin and the highest-enriched MSH4 gene hexile, CHG and CHH methylation showed a matching periodicity to the pattern of MSH4 coverage around the well-positioned nucleosomes (fig. 50B) (Stroud et al., 2013). This relationship was less clear for CG methylation, consistent with the more complex role of CG methylation associating with both repressed and active gene expression (Bewick & Schmitz, 2017; Zilberman, 2017). These relationships are particularly interesting given the characterised ten bp periodicities of DNA methylation in relation to nucleosomes (Chodavarapu et al., 2010). Therefore, these analyses indicate correlations between MSH4 binding and DNA methylation patterns in relation to nucleosomes.

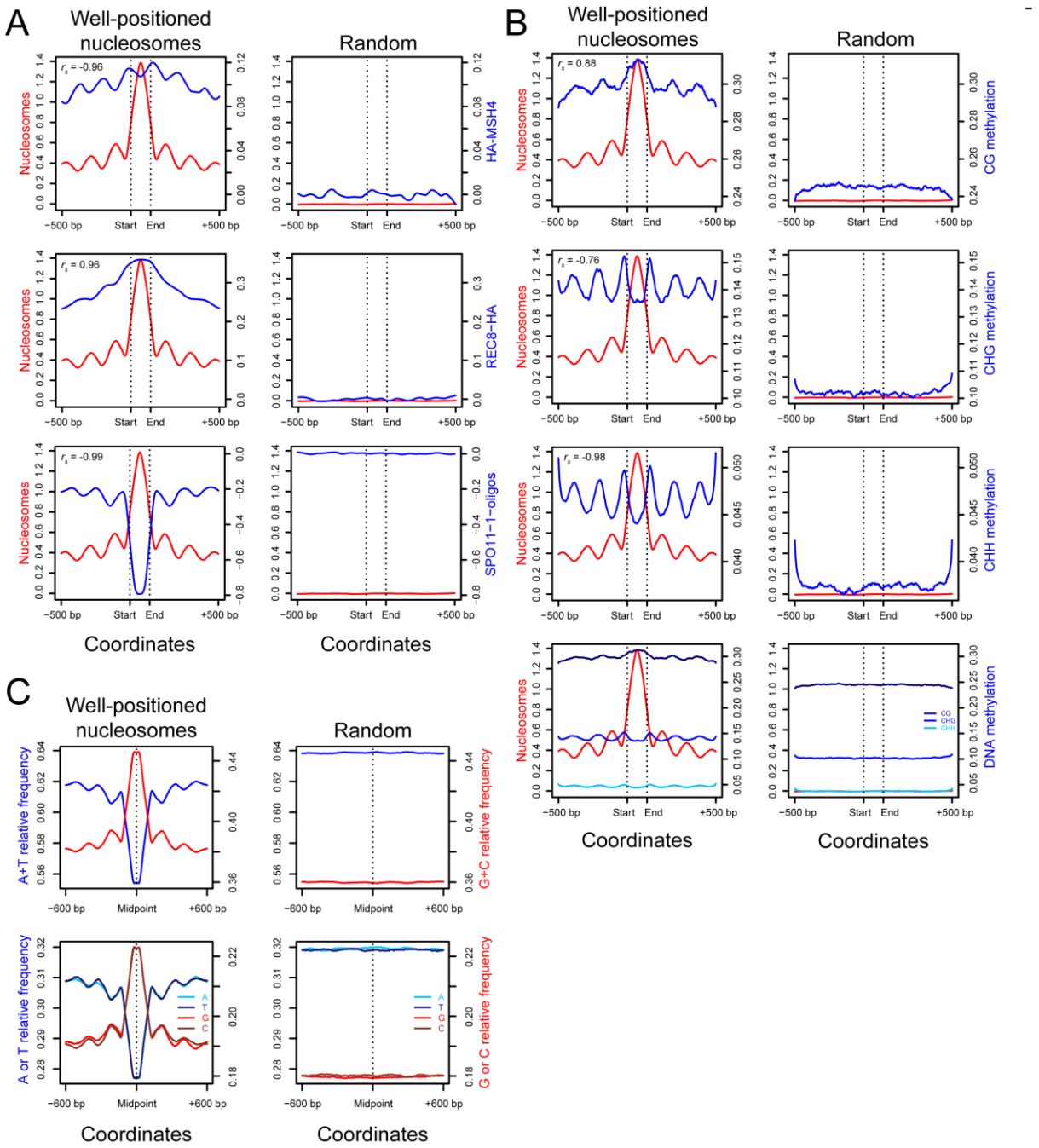


Figure 50. HA-MSH4 enrichment shows a putative anticorrelation with well-positioned nucleosomes at the fine-scale.

(A) Plots of MNase-seq (nucleosome) enrichment (red) at sites of well-positioned nucleosomes, including 500 bp flanking regions. This is compared with HA-MSH4 $\log_2$(ChIP/input) enrichment, REC8-HA ChIP-seq, and SPO11-1-oligos (blue) (Choi et al., 2018; Lambing et al., 2019). Spearman's correlation coefficients (r_S) are stated within the plot margins. (B) As for (A), but comparing MNase-seq (red) with methylation levels at CG, CHG and CHH contexts (blue).

5.14 Discussion

5.14.1 A consideration of MSH4 ChIP-seq enrichment in relation to immunocytological analyses.

ChIP-seq provides a complementary approach to studying meiotic recombination proteins, alongside cytological and genetic analyses. At the chromosome scale, this data presents a challenge to integrate with cytological studies of MutSγ distributions. In both Arabidopsis and rice, MSH4 and MSH5 form punctate foci across the chromosome arms, and there is no indication of specific enrichment over centromere proximal heterochromatin (Higgins et al., 2004, 2008b; Lu et al., 2008; Luo et al., 2013; Zhang et al., 2014). These studies, and those in other eukaryotes, led to an initial prediction that MSH4 ChIP-seq enrichment would be apparent throughout the chromosome arms, and depleted within the centromere. This was also intuited from the known role of ZMM proteins as pro-crossover factors, which would hence show a positive correlation with crossover frequency at the chromosome scale. The MSH4 ChIP-seq data in fact shows the inverse relationship, and is highly enriched over the crossover cold centromeres and is depleted around crossover midpoints at the fine-scale.

However, the nature of ChIP-seq data differs from cytological data in several key aspects. Firstly, it samples across a large population of meiocytes, at a scale not possible with immunocytology. In contrast, immunocytology captures the pattern of individual foci within individual cells. Therefore, ChIP-seq analyses may be able to detect underlying patterns below the detection level of cytological approaches. Secondly, ChIP-seq data has a wide dynamic range, and is able to detect very low levels of enrichment (Bailey et al., 2013). In contrast, cytological analyses must set thresholds for exposure time, and improve the definition and quality of images by removing background signal. However, this raises the question of to what extent these procedures may be obscuring biologically relevant signals that exist at lower levels of enrichment. Hence, it is possible that the centromeric association of the MSH4 ChIP-seq signal represents a ‘non-foci’ mode of binding – perhaps in association with the meiotic axis and cohesin superstructure – that is being missed with foci-based cytological analyses. Thirdly, it is unknown how such concentrated foci, such as those of MutSγ during leptotene, would translate into DNA sequencing molecules. For instance, if ten MutSγ heterodimers were in very close proximity – and associated into polymer-like structures such as those observed with other MMR-related proteins, such as the budding yeast Mlh3 (Manhart et al., 2017) – they may consequently only protect a single DNA fragment. This would become represented as a single mappable read, yet may be observed as a bright foci. This highlights the potential difficulty in unifying observations at the cytological scale, where signal intensity at a single site within a single nucleus can be captured, and ChIP-seq data, where this may not necessarily be the case.

A related concern is the potential bias against more complex DNA structures during sequence library preparation. As MutSγ is predicted to associate with early recombination intermediates, from single end invasions to HJs (Snowden et al., 2004, 2008; Lahiri et al., 2018), it will be important to consider how these structures would be affected by such procedures as crosslinking, end repair, adaptor ligation and size selection. In addition, the efficiency of crosslinking will vary throughout the genome, and one expects an increased crosslinking efficiency within the chromatin dense centromere-proximal regions because highly compacted chromatin has an increased chance of crosslinks forming between protein-protein and protein-DNA. This bias may in part explain the striking enrichment of MSH4 ChIP-seq at the centromeres and within heterochromatin. Assessing the possibility of this bias will prove an important line of inquiry, especially if it becomes possible to reverse the bias and actively enrich for more complex DNA substrates. However, assuming that the binding substrates of MutSγ are not being selected against during the ChIP-seq protocol, and that the *in vitro* binding assays translate into a biological context (Snowden et al., 2004, 2008; Lahiri et al., 2018), the data imply that the majority of the genome is competent for MSH4 binding and is hence able to form at least some types of meiotic recombination DNA intermediate.

However, there are several possible explanations for the elevated MSH4 enrichment over the pericentromere, which has also been observed in ChIP-seq experiments of SPO11-1 and REC8 (Lambing et al., 2019). Firstly, the ChIP-seq signal may be the result of sampling the entire time course of meiosis and, when averaging over thousands of cells, may be capturing the differential persistence times for different chromosome regions. For example, MutSγ heterodimers occurring in proximity to the centromere may be crossover suppressed, which leads to the stalling of recombination intermediates and hence an increased occupancy time for MutSγ. In contrast, the more crossover permissive chromosome arms may actually have a reduced occupancy time due to faster crossover maturation, i.e. a faster progression through the series of molecular events by which a crossover designated site become a *bona fide* crossover. However, to my knowledge there is no evidence supporting such a model. Secondly, whilst synapsis proceeds unimpeded in the Arabidopsis *msh5-1* mutant, with only a slightly increased duration, synapsis was significantly delayed in *msh4-1* (Higgins et al., 2004, 2008b). This potentially indicates a cryptic MSH5-independent role for MSH4 in promoting synapsis, which may explain the close association between MSH4 enrichment and markers of the chromosome axis and heterochromatin. This question could be directly addressed by performing a ChIP-seq analysis using a tagged MSH5 complementing line.

5.14.2 Interpretations of the MSH4 ChIP-seq enrichment profile in Arabidopsis.

The analyses presented in this chapter support some aspects of the tethered-loop/axis model. In this model, meiotic DSBs form within DNA sequences located in the chromatin loops, which extrude from the chromosome axis (Blat et al., 2002). However, DSB formation itself takes place in proximity of the chromosome axis, following chromatin loop tethering, where recombination then takes place (Blat et al., 2002). This model is consistent with the observations in budding and fission yeast that local Rec8 enrichment, at the base of the chromatin loops, supresses DSB formation (Storlazzi et al., 2010; Panizza et al., 2011; Nambiar & Smith, 2018), whilst in budding yeast Rec8-cohesin is required to promote DSB repair (Klein et al., 1999; Kim et al., 2010). Furthermore, Spo11-accessory proteins Mei4, Mer2, and Rec114 were observed to stably associate with the meiotic chromosome axis, implying that DSB formation depends on axial association (Panizza et al., 2011). In budding yeast, a member of the Set1 complex, the PHD finger domain protein Spp1, links DSB formation to H3K4me3 by binding this modification and simultaneously interacting with the Mer2 axis protein (Borde et al., 2009; Acquaviva et al., 2013; Sommermeyer et al., 2013; Cooper et al., 2016). The data presented in this chapter indicate that the recombination protein MSH4 is associating with the meiotic chromosome axis, i.e. with REC8-cohesin, and is negatively associating with SPO11-1-oligos. One interpretation is that after DSB induction, MSH4 is being loaded onto recombination intermediates in proximity with the chromosome axis, which may be consistent with the tethered-loop/axis model (Blat et al., 2002).

Given the increase in MSH4 enrichment towards the TTS, it is tempting to speculate that this distribution is being driven by transcription, which may drive the MutSγ heterodimers from 5' to 3' locations throughout the gene body. Consistent with this interpretation, the highest hexile of MSH4 enrichment was distinct in several respects. Firstly, MSH4 enrichment was elevated throughout the gene body, in contrast to the 5' to 3' bias observed for the other five hexiles. In this respect, MutSγ may be behaving in a manner analogous to somatic cohesin, which was shown in mouse to be pushed 5' to 3' through gene bodies by transcription (Busslinger et al., 2017). Specifically, ChIP-seq enrichment of somatic cohesin was shown to reduce within gene bodies as transcription was increased, an effect thought to be driven by the movement of Poll II-associated transcription bubbles (Busslinger et al., 2017). Secondly, the highest MSH4 hexile was also notably associated with negative regulators of transcription, such as the histone modifications H3K9me2, H3K27me1, H3K27me3, and histone variant H2A.W, as well as CHG and CHH methylation which mark constitutively silent genes (Stroud et al., 2013). Hence, various lines of evidence suggest transcription may be driving the distribution of MSH4 within gene bodies, although this hypothesis is based on correlation analyses and is lacking direct experimental support.

The possible role of transcription in distributing MutSγ across the gene body is also consistent with biochemical data indicating that, like other MutS MMR heterodimers (Jiang et al., 2005), MutSγ forms a sliding clamp which encircles one or more DNA duplexes (Snowden et al., 2004, 2008; Lahiri et al., 2018). Furthermore, the fine-scale analyses of MSH4 ChIP-seq enrichment in relation to WPNs argues that MutSγ is being restricted to inter-nucleosomal linker DNA, consistent with nucleosomes blocking the diffusion of MutSγ. Although this relationship may seem contradictory, given the consistently positive correlation between MSH4 enrichment and nucleosome occupancy at all other scales and contexts, I believe that this is due to a combination of factors. Firstly, this subtle relationship would likely be lost when assessing total nucleosome occupancy, due to the less fixed position of more labile nucleosomes. Secondly, it seems reasonable that this periodicity in relation to nucleosome occupancy would only be revealed at fine-scale, as such a relationship would be impossible to observe at the chromosome scale and improbable at the gene scale due to the normalisation processes essentially averaging the enrichment profiles across all genes, and hence losing the periodicity of nucleosome occupancy. Particularly interesting is the observation that this relationship is not observed for the REC8 ChIP-seq enrichment, which also shows a positive correlation with WPNs but no periodicity in its distribution over the flanking regions. Although this may be due to technically artefacts, such as REC8 having different crosslinking properties to MSH4, it is also possible that the increased size of REC8-cohesin (a ~30-50 nm ring complex) means that its association with nucleosomes is different to MSH4 (Zickler & Kleckner, 1999). Speculatively, this may be indicative of a higher order structure at sites of meiotic recombination, with REC8-cohesin associating with the meiotic chromosome axis and defining the chromatin loops, whilst at the fine-scale nucleosome occupancy is shaping the local distribution of MutSγ (Lambing et al., 2019).

The association between MSH4, nucleosomes and heterochromatin is particularly intriguing, given the assumed evolutionary origin of MutSγ from a MMR heterodimer (Culligan & Hays, 2000; Hays, 2002). Numerous studies have now shown that MMR components, including the MSH2-MSH6 heterodimer, are constitutively associated with nucleosomes within genes. For example, a ChIP analysis of MSH2 in budding yeast indicated that MSH2 was constitutively associated with genomic DNA at a low level (Evans et al., 2000). More recent studies have found evidence for a specific association between H3K36me3 and human MSH6 (Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009; Li et al., 2013; Huang et al., 2018) and the Arabidopsis MSH6 contains a Tudor domain which is known to bind to H3K4me3 (Zhao et al., 2018). Thus, the association between the MSH protein family and nucleosomes may be an evolutionarily conserved function, consistent with the Arabidopsis MutSγ also associating with regions of nucleosome density. Therefore, it is possible that MutSγ is associated with

chromatin throughout the genome, analogous to the association of MMR proteins with H3K36me3-modified nucleosomes prior to mismatch formation, and only forms concentrated foci when stimulated by local recombination activity.

At the fine-scale, MSH4 and REC8 are enriched within genes bodies, with greatest enrichment towards the 3′-end. This observation is similar with studies in budding yeast, which also observed that cohesin is pushed outwards from the 3′-end of genes (Pan et al., 2011; Sun et al., 2015). In both budding yeast and Arabidopsis, meiotic DSBs are concentrated in nucleosome free regions, such as immediately upstream of the TSS (Pan et al., 2011; Choi et al., 2018). However, fine-scale pollen-typing analysis at *RAC1* and *RPP13* indicated that crossovers are concentrated towards the 5′-end of gene bodies (Serra et al., 2018b), consistent with historical measurements of crossover frequency in Arabidopsis (Choi et al., 2013). Hence, the ChIP-seq data are consistent with a speculative interpretation that crossovers are concentrating at the 5′-end of gene bodies, at the interface between meiotic DSBs and components of the meiotic axis/recombination machinery.

5.14.3 Limitations of ChIP-seq to studying meiotic recombination proteins in Arabidopsis.

There are several limitations for using ChIP-seq to study meiotic recombination proteins. Firstly, our current inability to isolate and purify meiocytes from somatic bud tissues results in the majority of meiocytes being male in the collected samples. Whilst this is not a limitation for comparing this data to cytology studies, given that almost all studies in Arabidopsis derive meiocytes from dissected anthers, it is not possible to know whether the observations made at the fine-scale will hold true for female meiosis. Secondly, the ChIP-seq approach samples the entire time-course of meiotic recombination, and hence averages protein binding pattern(s) throughout this process. Furthermore, each stage progresses at a different rate, and longer stages are therefore likely to be over-represented in the dataset (Armstrong, 2013). As many meiotic proteins play dynamic roles, and the chromosomes show dynamic changes in organisation and compaction, this is also important to consider. Thirdly, given the highly dynamic nature of local chromatin landscapes, particularly during DNA repair and recombination transactions (Hauer & Gasser, 2017), caution must be made when interpreting correlations between ChIP-seq datasets. For example, overlapping enrichment does not necessarily imply co-localisation in any single biological context, as it is possible for two mutually exclusive proteins having an identical localisation pattern to appear positively correlated in ChIP-seq data sets. This is the result of data being averaged across thousands of individual cells. However, even taking these caveats into consideration, I believe the ChIP-seq approach is able to provide a powerful tool for addressing the general relationships between recombination proteins and the underlying chromatin landscape, despite not being able to define the more precise timings and relationships possible using other techniques. I

believe the ChIP-seq approach to studying meiotic recombination in Arabidopsis will develop in increasingly powerful directions, especially once an efficient protocol is developed for isolating meiocytes of a single stage.

5.14.4 Future experiments.

There are several next steps and future directions for this project. Firstly, it is important to replicate the HA-MSH4 ChIP-seq result, to validate the relationships observed with the first biological replicate. In addition to the complementation analyses, an important future experiment is to validate the behaviour of the HA-MSH4 protein via immunocytology. It will be important to confirm that the tagged protein is distributing as punctate foci consistent with reports using antibodies raised against the endogenous protein (Higgins et al., 2004, 2008b). For example, it is possible that this analysis, using a high affinity antibody-HA-tag combination will reveal a previously hidden association between MSH4 and the centromere. Because the *HA-MSH4 msh4-1* line has been combined with a series of meiotic mutants (including *spo11-1-3*, *mer3-1*, *rec8-1*, *msh5-1*, and *hei10-2*), these lines will provide a powerful resource for dissecting the relationship between Arabidopsis ZMM proteins and MSH4 localisation (Grelon et al., 2001; Cai et al., 2003; Mercier et al., 2005; Higgins et al., 2008b; Chelysheva et al., 2012).

The close correlation between MSH4 and REC8 ChIP-seq enrichment across varying physical scales was unexpected, given the different biological roles of these proteins. An important next step will be to validate the implied association between these proteins. For instance, MSH4 enrichment over the centromere could be confirmed by immunolocalising the protein in meiotic chromosome spreads, whilst also using fluorescence *in situ* hybridisation techniques to mark the centromeres (Lambing et al., 2019). The ChIP-seq analysis also predicts a high degree of co-localisation between MSH4 and REC8, which could be validated through co-localisation studies in the *HA-MSH4 msh4-1 REC8-Myc rec8* line (Lambing et al., 2019). This line could also be used for biochemical studies, such as pull down assays to assess whether the proteins are physically interacting *in vivo*. Notably, co-immunolocalisation studies using rice pollen mother cells found that MSH5 and REC8 foci had a high degree of overlap at the leptotene stage, when REC8 also forms a more punctate pattern prior to its continuous localisation across pachytene chromosomes (Zhang et al., 2014).

Finally, the introduction of a *HEI10* overexpressor transgene into the *HA-MSH4 msh4-1* background makes it possible to perform a HA-MSH4 ChIP-seq experiment in this background, to assess the distribution of MSH4 in a background with highly elevated crossover frequency throughout the chromosome arms (Ziolkowski et al., 2017). Given that HEI10 is a predicted regulator of MSH4 stability, I predict a changed distribution in the background at the

chromosome and fine-scale, and such changes would in turn support the validity of the wild type *HA-MSH4 msh4-1* ChIP-seq result. I am excited to see the results from these experiments in the near future.

5.15 Acknowledgements

I would like to thank Dr Andrew Tock for performing the bioinformatic analyses of the MSH4 ChIP-seq data. Dr Christophe Lambing and Dr Kyuha Choi helped to pioneer the successful application of ChIP-seq to studying meiotic proteins in Arabidopsis. I would also like to thank Dr Sandra Cortijo for kindly sharing her Arabidopsis ChIP protocol, as well as her various ChIP-seq tips and tricks.

Chapter six – General discussion

The findings presented in chapters three and four raise several important questions. Firstly, how to unify the observed genome-wide positive correlation between interhomolog polymorphism and crossovers with the well characterised role of polymorphism in repressing crossover formation at the fine-scale. Secondly, how the observed *msh2-1* mutant crossover phenotypes in Arabidopsis can be reconciled with contrasting observations in other model organisms. These questions will now be addressed sequentially.

6.1 A unified model for the role of interhomolog polymorphism in shaping meiotic crossover distributions.

At the fine-scale, interhomolog polymorphism is a potent inhibitor of meiotic crossover formation, as demonstrated in numerous model systems (Dooner, 1986; Borts & Haber, 1987; Jeffreys & Neumann, 2005; Baudat & de Massy, 2007; Cole et al., 2010; Serra et al., 2018b). In Arabidopsis, pollen typing at two resistance gene crossover hotspots, *RAC1* and *RPP13*, revealed a strong negative correlation with interhomolog polymorphism within the analysed regions (Spearman's r = -0.69 and -0.89, respectively) (Serra et al., 2018b). This is consistent with reports in mouse, where sperm typing at the *A3* crossover hotspot revealed a crossover refractory region associated with an indel polymorphism (Cole et al., 2010). At the maize *bronze* (*bz*) locus, mapping of meiotic crossover junctions, using two different hybrid combinations (McC×W22 & McC×B73), found that crossovers negatively correlated with polymorphism across the 6.7 kb *bz* interval (Dooner & He, 2008). Therefore, assaying crossover positions across endogenous genetic intervals clearly demonstrates that interhomolog polymorphism acts to repress local crossover frequency.

However, meiotic recombination between polymorphic substrates appears less repressed in comparison to somatic HR, where even a single mismatch causes potent suppression of HR in budding yeast and Arabidopsis (Datta et al., 1997; Li et al., 2006). A close inspection of crossover distributions at hotspots shows that, although negatively correlating with polymorphism, crossovers are still able to occur in relatively SNP-dense regions (Serra et al., 2018b). Indeed, the technique of pollen typing depends on a relatively high polymorphism density in order to map crossover positions across a defined interval (Choi et al., 2013; Drouaud et al., 2013; Serra et al., 2018b). Taken together, these studies across diverse eukaryotes indicate that whilst at any single locus interhomolog polymorphism appears to negatively correlate with crossover frequency, it is not an absolute inhibitor of crossover

occurrence, and in many cases total interhomolog polymorphism density does not correlate with total genetic distance across a given interval (Hilliker et al., 1991; Cole et al., 2010). Hence, meiotic crossovers appear to escape the stringent repression of recombination between divergent sequences occurring in somatic cells.

The fine-scale repressive effects of interhomolog polymorphism on crossovers would therefore predict a negative correlation with diversity genome-wide. The fact that in wild type Arabidopsis hybrids the opposite relationship is observed raises the question of how polymorphism relates to the hierarchy of crossover regulation at different scales, and initially appears contradictory. However, I believe that the fine-scale and chromosome-scale observations are not incompatible for several reasons. Firstly, although many investigations of crossover frequency at the fine-scale observe a negative relationship, many of these genetic intervals are located in regions of high sequence diversity relative to the genome-average, such as *RAC1* and *RPP13* (Serra et al., 2018b). For example, the experiments at the *A3* crossover hotspot in mouse used hybrids with 32 polymorphisms within the 2 kb sequence (1.6% divergence) (Cole et al., 2010). Furthermore, *Mus musculus* × *Mus musculus spretus* hybrids showed an even higher crossover frequency, despite a higher degree of polymorphism (2.4% divergence) (Cole et al., 2010). Therefore, despite having a negative relationship with crossover distributions at the fine-scale, SNP density is a poor predictor of crossover frequency at the scale of hotspots (several kb). However, it is also unclear whether this is due to the absence of a relationship with polymorphism at a broader scale or whether this relationship is being overshadowed by other *trans* or *cis* factors, such as dosage dependent crossover modifiers or epigenetic modifications to the local chromatin environment, respectively (Yelina et al., 2015; Ziolkowski et al., 2017).

A second consideration is that one would not necessarily expect the same factors to determine the fine-scale and chromosome-scale distribution of crossovers. For instance, many parameters known to influence meiotic recombination are non-randomly distributed across the genome. For example, meiotic DSBs are concentrated in the chromosome arms and regions of low nucleosome density, and the meiotic recombination proteins MSH4 and REC8 show variable patterns across the genome, with both proteins showing striking enrichment over the pericentromeres and centromeres (see section 5.6) (Choi et al., 2018; Lambing et al., 2019). Therefore, there seems to be no reason to believe that a fine-scale negative correlation with polymorphism should imply a negative correlation at the Megabase-scale. Indeed, although at the chromosome-scale crossovers positively correlate with interhomolog polymorphism, this association may be entirely a consequence of other causal factors which also correlate with polymorphism density. However, the data presented in this study indicate that an interaction

between MMR and interhomolog polymorphism does contribute to shaping the genome-wide crossover landscape, although in a positive way that was counter to my initial expectation.

In the context of observations demonstrating a clear negative correlation between interhomolog polymorphism and crossover frequency within crossover hotspots, it was particularly interesting and unexpected to find a positive association with GBS crossover midpoints and SNPs (Serra et al., 2018b). However, I propose that there are several explanations for this apparent discrepancy. Firstly, in contrast to the highly accurate and deep measurements of crossover distributions across a single locus using the pollen typing technique, GBS crossovers are assigned within an average interval size of ~1 kb, as determined by crossover positions and inter-SNP distances. However, there is also some uncertainty in the GBS crossover positions, as demonstrated by attempts to validate GBS crossover junctions using Sanger sequencing or a comparison of crossover positions assigned from identical GBS libraries sequenced to high or low depth (discussed in section 4.2) (Rowan et al., 2015; Serra et al., 2018a). These investigations revealed that 13% of GBS crossover midpoints were called within identical SNP intervals, and that the median difference between the low and high sequencing crossover sites was 1,137 bp (Serra et al., 2018a). This distance is consistent with estimates of GBS accuracy based on Sanger sequencing (Rowan et al., 2015). Hence, the power of GBS to resolve crossover locations may be in the range of 1-2 kb on average, beneath the level of precision required to observe the fine-scale relationships with polymorphism observed within crossover hotspots (Serra et al., 2018b). Therefore, the analysis of SNP distributions within 10 kb windows centred on GBS crossover midpoints may be revealing a positive association with interhomolog polymorphism at a scale intermediate to the genome-wide crossover distributions and the fine-scale. However, it still remains a concern that the local association between GBS crossovers and SNPs is an artefact of the TIGER analysis pipeline, which may bias crossovers to be called in regions of elevated SNP density (discussed in section 4.10).

However, consistent with GBS crossovers capturing a positive intermediate-scale association with polymorphisms, and also with a role for MMR in determining the chromosome-scale positive association between crossovers and polymorphism, the Col/Ler and Col/CLC *msh2-1* mutants were the only genotypes where this positive association in proximity to crossover midpoints was reduced. This was most clear when assessing differences in SNP distribution surrounding the *msh2-1* Col/Ler hybrid crossovers or a random simulation. In contrast to the Col/Ler wild type, the pattern of statistical significance across the 10 kb window was comparable to that observed for comparisons between pairs of random simulations, although it was notable that the significance threshold was still crossed at the position of the crossover midpoints. Therefore, it is tempting to speculate that a similar process is driving both the

megabase-scale association between crossovers and SNPs as well as the positive association observed across the 10 kb windows, at a hierarchical level above the fine-scale positioning of crossovers in relation to polymorphism (i.e. within hotspots).

6.2 Possible models for the role of MMR in shaping meiotic recombination landscapes in Arabidopsis.

The crossover redistribution observed in this investigation implicates MMR activity in mediating a positive association between crossovers and interhomolog polymorphism genome-wide. This raises the question of what mechanism(s) might be causing this effect, given the well characterised antirecombination activity of MMR during somatic recombination between divergent repeat sequences in a wide range of eukaryotes (de Wind et al., 1995; Elliott et al., 1998; Elliott & Jasin, 2001; Li et al., 2006; Welz-Voegele & Jinks-Robertson, 2008). One tempting explanation is derived from observations of regional biases in MMR activity on mismatches caused by somatic DNA replication errors (Frigola et al., 2017; Belfield et al., 2018). In Arabidopsis, the *msh2-1* mutation increases the mutation rate within genes to a greater extent than centromere proximal regions, when compared to the mutation rate in wild type (Belfield et al., 2018). This observation is consistent with the reduced mutation rate within exons, as computed from ENCODE cancer cell line data, which again suggests the activity of MMR is biased across the genome (Frigola et al., 2017). Mechanistically, several studies have demonstrated an association between human MSH6 and the histone modification H3K36me3, a modification enriched over gene bodies, which suggests that MMR proteins may be recruited to gene sequences by specific chromatin modifications (Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009; Li et al., 2013; Huang et al., 2018). Although the PWWP domain responsible for H3K36me3-association is not present in the Arabidopsis MSH6 ortholog, it does have a Tudor domain promoting an interaction with H3K4me3 (Zhao et al., 2018). Notably, H3K4me3 is also a marker of actively transcribed genes and is located at the 5'-end of gene bodies (Zhang et al., 2009; Choi et al., 2013). Together, these studies suggest that MSH2-heterodimers associate with chromatin, prior to the occurrence of a DNA mismatch, and that they are concentrated in proximity to genes.

Therefore, it is possible that the genome-wide positive association between crossovers and interhomolog polymorphism is in fact being driven by increased crossover repression in the gene-rich chromosome arms, due to the preferential localisation of MMR proteins in these regions, which would cause a relative increase in centromere proximal crossovers. This is because crossover progression in the pericentromeres would be less impeded by MMR, and crossover designation in these regions would cause a repression of crossovers in the adjacent chromosome arms due to crossover interference (Copenhaver et al., 2002). If this is the case,

I would predict that crossover landscapes within inbred Arabidopsis individuals may show a distalisation relative to the hybrids, although this would be difficult to assess given the requirement for genetic markers in mapping crossover positions.

Moreover, regional differences in MSH2-heterodimer enrichment due to the chromatin landscape may also explain the chromosome-scale crossover redistribution away from the centromere and into regions of reduced polymorphism density observed in in the Arabidopsis *msh2-1* mutant, relative to the wild type. This may occur because a loss of MMR activity would cause a relatively greater reduction in MMR-mediated anti-recombination within the gene-rich chromosome arms, due to the recruitment of MutSα heterodimers via gene-associated marks including H3K4me3. Correspondingly, crossover interference from the increased crossovers in the chromosome arms would cause a relative reduction in crossovers close to the centromeres. This effect would be enhanced due to the smaller effect of losing MMR anti-recombination close to the centromeres, as MMR plays a less active role in these regions (Belfield et al., 2018). Consistent with this model, and the presence of a H3K4me3-interacting Tudor domain in MSH6, crossovers in the *msh2-1* mutant were located in regions of elevated H3K4me3 enrichment, compared to the wild type (Zhao et al., 2018).

A second potential model to explain the crossover redistribution effect in the *msh2-1* mutant, as well as the positive association with meiotic crossovers and diversity genome-wide, was previously proposed to explain the *cis* effect (Ziolkowski et al., 2015). The authors hypothesised that crossover precursors containing mismatches would be delayed in maturation, which would trigger feedback mechanisms leading to additional meiotic DSBs (Ziolkowski et al., 2015). Such meiotic DSB feedback has been previously demonstrated in several eukaryotes, including Arabidopsis and budding yeast (Carballo et al., 2008; Lange et al., 2011; Zhang et al., 2011; Kurzbauer et al., 2012; Garcia et al., 2015). For example, the signal transduction kinases Mec1 (ATR) and Tel1 (ATM) act in *trans* to inhibit the formation of DSBs at equivalent positions on both homologous chromosome pairs (Zhang et al., 2011), with Tel1 additionally suppressing DSB formation adjacent to existing DSBs on the same chromosome (in *cis*) (Garcia et al., 2015). Therefore as a consequence of delayed crossover progression, these 'late' DSBs may produce more crossover precursors and ultimately a higher crossover frequency in regions of relatively higher interhomolog polymorphism. These crossovers would in turn reduce crossover frequency in adjacent sequences due to interference, which acts at the megabase-scale, causing a reduction in crossover frequency in less polymorphic sequences (Ziolkowski et al., 2015). Support for this model is provided by a study in budding yeast, which showed that DSB density is elevated in the *zip1*, *zip3* and *msh5* backgrounds, indicating that meiotic recombination suppresses additional DSB formation (Thacker et al., 2014). Assuming that delayed crossover maturation at mismatched

intermediates depends on their identification by MSH2-heterodimers, this could explain the reduced association between crossovers and interhomolog polymorphism in the *msh2-1* mutant. Experimental validation of this feedback hypothesis should be possible by mapping SPO11-1-oligonucleotides in Col/Ler hybrids, which this model would predict to be elevated in regions of high interhomolog polymorphism, relative to the inbred (Choi et al., 2018; Underwood et al., 2018). An alternative, but related, model was also proposed, whereby crossovers occurring at regions of high interhomolog polymorphism are more likely to be Class I interfering crossovers, as evidenced by the loss of Class II crossover increases in *fancm* mutant hybrids, relative to inbreds, and the highly distalised crossover increases in *recq4a recq4b* double mutants (Giraut et al., 2015; Fernandes et al., 2018a; Serra et al., 2018a).

A further intriguing hypothesis, not mutually exclusive with the previous models, is the overlap in components between the MMR machinery and the Class I crossover pathway machinery (Cooper et al., 2018). Specifically, the MLH1-MLH3 heterodimers mediate both mismatch processing and Class I crossover resolution via their endonuclease activity (Rogacheva et al., 2014). Hence, the stimulation of MMR caused by mismatches occurring during meiotic recombination, although causing local repression, may counterintuitively stimulate class I crossovers at recombination intermediates in proximity to the mismatch. In the context of Arabidopsis hybrids, which show Megabase-scale variation in interhomolog polymorphism density, MLH1-MLH3 heterodimers may be preferentially concentrated in regions of greater interhomolog polymorphism, thereby stimulating crossovers within these regions. The loss of MSH2 activity would therefore cause a redistribution of crossovers away from divergent sequences, and hence away from centromere-proximal regions.

In summary, I propose that meiotic crossovers in Arabidopsis are positively associating with interhomolog polymorphism, at the chromosome-scale, due to either a) relative enrichment of MSH2-heterodimers within genes causing a greater crossover repression in the chromosome arms, and hence causing a relative crossover increase in gene-poor pericentromeres, b) the recruitment of Class I crossover components to sites of MMR, which occur disproportionately close to the divergent centromere and indirectly cause the occurrence of Class I crossovers in these regions, or c) the elevated activity of MMR in regions of increased polymorphism density leading to a delay in recombination progression, and resulting in a wave of crossover-promoting ‘late’ DSBs (fig. 51) (Ziolkowski et al., 2015; Cooper et al., 2018). Although these hypotheses are not mutually exclusive, establishing how the balance of direct anti-crossover versus indirect pro-crossover activity is established in different genomic contexts is hard to predict. One may expect that the influence of MMR on meiotic recombination outcomes varies greatly depending on total interhomolog polymorphism density, as well as its relative distribution across the genome.

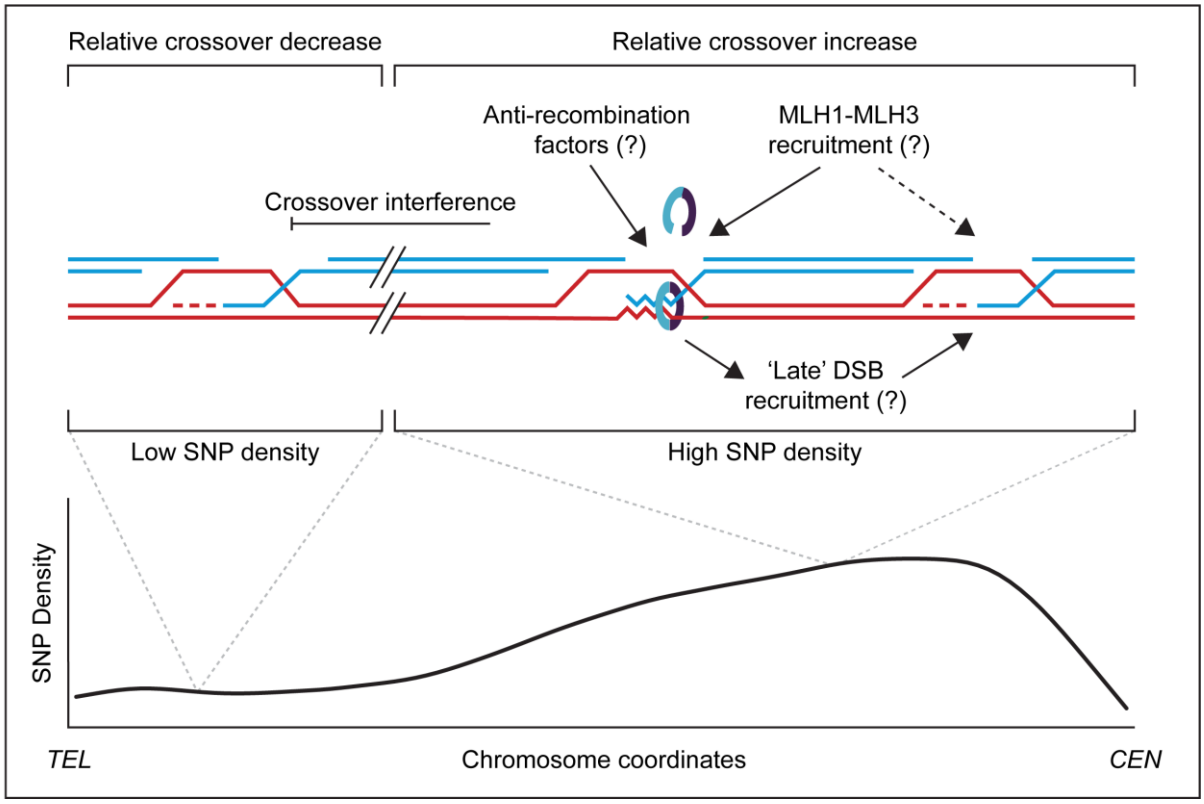


Figure 51. A model of MSH2 activity during meiotic recombination in Arabidopsis.

In regions of low SNP density (left), strand invasion has a reduced probability of generating mismatches as a consequence of interhomolog polymorphism. In regions of high SNP density (right), MSH2-heterodimers (pale blue, purple) target mismatches occurring within strand invasion intermediates, recruit anti-crossover factors (grey, green, yellow), and hence repress crossover formation. However, through unknown mechanisms (potentially via recruitment of MLH1-MLH3 heterodimers or 'late' DSBs) MMR activity promotes meiotic crossovers at a mega-base scale, which potentially suppress crossovers in adjacent regions due to crossover interference.

6.3 Reconciling the observations and models of MMR activity in Arabidopsis meiosis with observations in other eukaryotic systems.

The characterisation of Arabidopsis *msh2-1* mutant phenotypes presented here extends previous findings that putatively demonstrated MSH2 to act as a potent hybrid-specific anti-crossover factor (Emmanuel et al., 2006). This present study reveals that Arabidopsis hybrids with an average sequence divergence of ~0.55% show no reduction in fertility, whilst fertility is slightly reduced in the *msh2-1* mutant. However, this reduction was comparable to that observed in *msh2-1* inbred backgrounds, indicating that this effect was likely due to DNA mutation accumulation in *msh2-1*. Interestingly, two studies of *msh2* S288C/SK1 budding yeast hybrids, which are ~0.57% diverged, observed that spore viability decreased by 10-21% in the *msh2* mutant (Martini et al., 2011; Cooper et al., 2018). Hence, these results are also consistent with the exposure of deleterious mutations in haploid spores. However, in contrast to the Arabidopsis hybrid tested (Col/CLC), the budding yeast hybrid showed a significant reduction (8-30%) in spore viability in the wild type hybrid, relative to the inbred strains (Martini et al., 2011; Cooper et al., 2018), indicating a higher degree of hybrid incompatibility.

This latter observation perhaps explains the reason for a 26-40% increase in total crossover number in the S288C/SK1 budding yeast *msh2* hybrid (Martini et al., 2011; Cooper et al., 2018), and implies that budding yeast hybrids may be more sensitive to crossover suppression via meiotic MMR activity. It is therefore possible that a similar phenotype would manifest if using more divergent Arabidopsis intra- or inter-species hybrids, such as those observed with yeast interspecies hybrids (Hunter et al., 1996). The data presented in the current study also differ with respect to reports of *msh2* phenotypes in tomato, which showed a severe disruption of meiotic recombination and near sterility in *msh2* inbred mutants (Sarma et al., 2018). However, this more extreme phenotype may be due to the more complex and repetitive nature of the tomato genome, in contrast to Arabidopsis, which also has a large quantity of heterochromatin (Gouil & Baulcombe, 2016). In this genomic context, I speculate that MMR activity may play an essential role in the fidelity of the meiotic recombination process, and repress ectopic recombination between homologs.

Notably, a recent study calculated the SNP density surrounding crossover sites in wild type or *msh2* S288C/SK1 budding yeast hybrids strains, in a 1 kb window (Cooper et al., 2018). In contrast to the current study, the authors observed that wild type crossovers were biased towards SNP depleted regions, relative to the expected distribution based on the meiotic DSB landscape (5.32 vs. 6.18 SNPs/kb) (Cooper et al., 2018). Again in contrast to the current study, they observed that *msh2* crossovers occurred in regions of elevated SNP density, relative to the wild type, and were not significantly different from the random expectation (6.26 vs. 6.18

SNPs/kb) (Cooper et al., 2018). This result was validated using an additional independent dataset, in the S96/YJM789 hybrid, and the effect was diminished with increasing distance from the crossover site, as assessed using 2 or 4 kb windows (Cooper et al., 2018). This indicated that interhomolog polymorphism represses crossover formation, consistent with crossover hotspot analysis in Arabidopsis (Serra et al., 2018b). Interestingly, the authors also observed that Class I DNA repair pathway mutants phenocopied the *msh2* mutant, implying that enrichment of wild type crossovers in relatively SNP depleted regions was dependent on this pathway (Cooper et al., 2018). This led the authors to propose that MMR at sites of interhomolog mismatches may be recruiting DNA repair factors that also play a role in resolving Class I crossovers, such as MLH1 and MLH3 (Cooper et al., 2018). Their observation implies that Class II non-interfering crossovers are relatively insensitive to interhomolog polymorphism, which is again in striking opposition to Arabidopsis (Girard et al., 2015; Ziolkowski et al., 2015; Fernandes et al., 2018a; Serra et al., 2018a).

These observations raise the question of why the *msh2* crossover phenotype is so different between budding yeast and Arabidopsis. Due to the deep evolutionary conservation of the MMR machinery in eukaryotes (Kunkel & Erie, 2015), it seems unlikely that either species represents an anomaly as a consequence of differences in MMR activity, which implies that differences in the genome architecture or recombination machinery have led to differing *msh2* phenotypes. Therefore, these contrasting results also provide a useful criteria for the model in Arabidopsis, which must be capable of explaining the contrasting phenotypes in budding yeast.

Notably, the budding yeast genome is 12.1 Mb whilst the Arabidopsis genome is 119.1 Mb. Furthermore, budding yeast have 16 chromosomes, approximately three times that of Arabidopsis. Therefore, the average budding yeast chromosome is 0.754 Mb, whilst that of Arabidopsis is 23.8 Mb. Moreover, ~74 crossovers occur per meiosis in S288C/SK1 wild type hybrids, compared to ~10 crossovers per meiosis observed for Col/Ler hybrids (Wijnker et al., 2013; Cooper et al., 2018). This implies that meiotic recombination is occurring on a vastly different scale in Arabidopsis, with a far smaller proportion of the genome involved in crossover recombination (fig. 52A,B). However, it is unclear how these differences would themselves alter the relationship between polymorphism, MMR and crossover outcomes. One explanation may be the ratio of MSH2-heterodimers to DNA content, which may determine the efficacy of MMR to act upon mismatched bases within meiotic recombination structures. Unless Arabidopsis meiocytes contained 10-fold the quantity of MSH2-heterodimers and MMR proteins, one might expect the efficacy of MMR-mediated anti-recombination to be reduced within this system.

However, there are also striking differences in relation to the interhomolog polymorphism landscapes of Arabidopsis hybrids and S288C/SK1 hybrids. In Arabidopsis Col/Ler hybrids, polymorphism density varies dramatically across the genome, and shows pronounced variation at megabase-scale. For example, SNP densities are elevated in proximity to the centromeres, and diversity is thought to be highest within the centromeres themselves, although analysis of these regions is hindered due to a lack of genome assembly (fig. 52C) (Simoens et al., 1988; Hosouchi et al., 2002; Borevitz et al., 2007; Ito et al., 2007). In contrast, the density of interhomolog polymorphism in the S288C/SK1 hybrid is far more homogenous, and does not show a clear spatial relationship with particular genomic features, such as the centromeres (fig. 52C). This difference may be crucial in understanding the differing fertility phenotypes of *msh2*. For example, given such striking variation in polymorphism density across the Arabidopsis genome, there is a high probability of many recombination events experiencing no or relatively few mismatches. This variation may be the reason why Arabidopsis Col/CLC hybrids do not have fertility defects, despite having an average interhomolog polymorphism density comparable to that of S288C/SK1 hybrids.

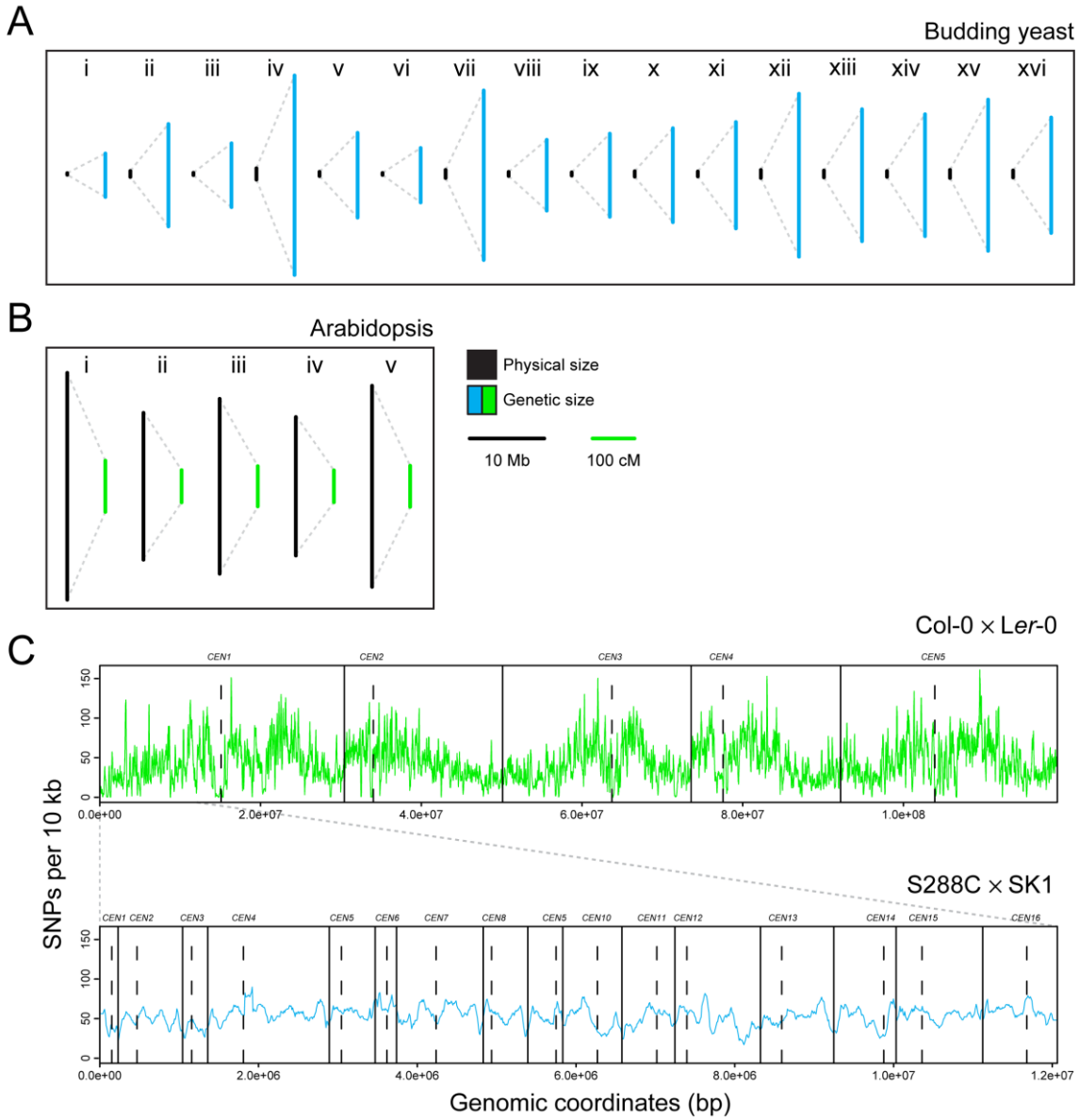


Figure 52. A comparison of the budding yeast and Arabidopsis physical and genetic maps, and the genome-wide SNP distribution in the Arabidopsis Col/Ler and budding yeast S288C/SK1 hybrids.

(A) A scaled representation of the budding yeast (S288C) physical size (black line) and genetic size (pale blue line), for chromosomes i-xvi. (B) As for (A), but showing the Arabidopsis physical size (black line) and genetic size (green line) drawn to the same scale as (A), for chromosomes 1-5. The budding yeast genetic map is from Cherry et al. (1997), and the Arabidopsis genetic map is from Meinke et al. (2009). (C) SNP density is plotted in 10 kb sliding windows for Col/Ler hybrid (green line) and the S288C/SK1 hybrid (blue line). The positions of centromeres (*CEN1–CEN5* or *CEN1–CEN16*) are shown with vertical dashed lines and telomeres with vertical straight lines. The grey lines connecting the plots indicate the physical scale of the budding yeast genome in relation to the Arabidopsis genome. Col/Ler SNP densities were calculated from the 1,135 genomes datasets (Alonso-Blanco et al., 2016), and S288C/SK1 SNPs were kindly shared by Dr Timothy Cooper (Cooper et al., 2018).

Furthermore, megabase-scale regional differences in polymorphism density in Arabidopsis may explain the differing *msh2* crossover phenotypes, as they create the possibility for feedback processes to emerge within and between different genomic regions. For example, in Arabidopsis, SNP dense regions may begin to sequester the MMR factors MLH1 and MLH3, due to the frequent occurrence of mismatches, which may indirectly lead to increased Class I crossovers within these regions, and thus cause a positive association between crossovers and interhomolog polymorphism at the chromosome-scale (Jackson et al., 2006; Dion et al., 2007; Cooper et al., 2018). In support of this, *in vitro* data indicate that the budding yeast Mlh1-Mlh3 endonuclease is stimulated by the Msh2-Msh3 heterodimer at sites of mismatched bases, and Mlh1-Mlh3 may therefore concentrate in regions of denser interhomolog polymorphism during meiosis (Rogacheva et al., 2014). Crossovers would then be supressed in adjacent regions, due to the effect of crossover interference, causing a repression of crossovers within the less divergent chromosome arms (Copenhaver et al, 2002). Such a phenomenon would not occur in budding yeast hybrids, due to the more homogenous distribution of interhomolog polymorphism (Cooper et al., 2018). In contrast to Arabidopsis, the 'late' DSB feedback model would not predict a redistribution of crossovers in the S288C/SK1 hybrid, into regions of relatively higher interhomolog polymorphism, as this model requires the presence of megabase-scale variation in polymorphism density (Ziolkowski et al., 2015).

A further important difference between Arabidopsis and budding yeast is the extent and organisation of chromatin. Gene density is relatively high across the budding yeast genome, which lacks megabase-scale regions of gene depletion, as seen in Arabidopsis (Fig. 42B) (AGI, 2000; Duina et al., 2014). In contrast with Arabidopsis, which shows high levels of DNA methylation, particularly in the pericentromeres and centromeres, budding yeast lacks DNA methylation (Proffitt et al., 1984; Stroud et al., 2013). Furthermore, budding yeast is also unusual in lacking the heterochromatic histone modifications H3K9me2 and H3K9me3, which are found in most other eukaryotes (Bühler & Gasser, 2009). Instead, budding yeast silences chromatin via a histone deacetylation pathway, mediated by the SIR proteins (Bühler & Gasser, 2009). Electron cryotomography was used to investigate the higher order chromatin structure in budding yeast, which revealed no evidence for chromatin structures larger than tetra-nucleosomes (Chen et al., 2016). The budding yeast chromatin state was found to remain accessible throughout both interphase and mitosis, consistent with high levels of transcription (Chen et al., 2016). Budding yeast therefore lacks the large, epigenetically defined centromeres found in many eukaryotes, including Arabidopsis, and forms ~125 bp 'point' centromeres (Henikoff & Henikoff, 2012). High resolution MNase-seq showed these centromeres to comprise single CenH3-variant-containing nucleosomes (Henikoff & Henikoff, 2012), which is in striking contrast to the megabase-scale, highly heterochromatic and

repetitive centromeres in Arabidopsis (Simoens et al., 1988; Hosouchi et al., 2002; Fransz et al., 2006; Borevitz et al., 2007; Ito et al., 2007).

The striking differences between Arabidopsis and budding yeast chromatin landscapes may explain the differences in *msh2* crossover phenotypes. For instance, if MSH2-heterodimers are pre-associated with sites of active, open chromatin, this would have little effect on their distribution across budding yeast meiotic chromosomes, which are constituted primarily of open chromatin, albeit in the context a meiotic chromosome axis (Chen et al., 2016; Schalbetter et al., 2018). In Arabidopsis, an association between MMR components and open chromatin (e.g. H3K4me3) would lead to variation in MMR activity at the megabase-scale, as evidenced by relatively higher mutation rates close to the centromeres (Ossowski et al., 2010; Belfield et al., 2018). In summary, these key differences in both the form and scale of chromatin landscapes between Arabidopsis and budding yeast may be key to interpreting the different roles of MMR in regulating meiotic crossovers.

6.4 Perspectives and future experiments.

The data presented in this study have generated several possible models for the role of MMR in meiotic recombination in Arabidopsis, and have revealed several surprising phenotypes in the *msh2-1* mutant. An important experiment to distinguish between the two proposed, not mutually exclusive, models is to test whether the *cis* effect is still present in an MMR-deficient background. Due to the reduced positive association between crossovers and genome-wide polymorphism density in *msh2-1*, I predict that the *cis* effect will be reduced or absent in *msh2-1*. In addition to using fluorescent crossover reporters to study the *cis* effect, it will also be important to use these reporters to investigate the impact of *msh2-1* on crossover interference.

Although several studies in budding yeast have monitored the effects of polymorphism on crossover frequency within genetic intervals in either wild type or MMR-deficient backgrounds (Borts et al., 1990; Chen & Jinks-Robertson, 1999), to my knowledge no study has monitored the fine-scale distribution of crossovers across a single genetic locus in a *msh2* mutant, at high depth and resolution. Validating the putative role of MMR in driving the fine-scale negative correlation between interhomolog polymorphism and crossovers at Arabidopsis crossover hotspots will be a priority.

Several experiments could address the proposed regional distribution of MMR proteins across the genome. Firstly, the construction of Arabidopsis lines expressing a tagged MSH2 protein will enable direct observations of MSH2-heterodimer localisation across meiotic chromosomes, which could be co-localised with meiotic recombination proteins, such as RAD51. Complementation could be assessed by passaging the *msh2-1* mutant with and without a *MSH2* transgene, to assess if the production of mutant phenotypes is reduced in the

transgenic lines (Leonard et al., 2003). Assembling a *MSH2* construct driven by a meiosis-specific promoter, such as the REC8 promoter (Lambing et al., 2019), will also enable the application of a meiotic ChIP-seq protocol to the study of this protein's genome-wide distribution, in both inbred and hybrid backgrounds. For example, one might expect MSH2 ChIP-seq occupancy to increase in regions of greatest heterozygosity during meiosis. However, it would also be informative to study the distribution of MSH2 on somatic chromosomes, as evidence suggests that the budding yeast and Arabidopsis epigenomic landscape is broadly conserved between mitosis and meiosis (Zhang et al., 2011b; Choi et al., 2018; Underwood et al., 2018). For example, correlations are observed between histone modifications measured in somatic cells and meiotic recombination markers, such as SPO11-1-oligos (Choi et al., 2018; Underwood et al., 2018), and budding yeast histone modifications and nucleosome positioning were relatively unchanged through meiosis (Zhang et al., 2011b). Therefore, studying the distribution of MSH2 across somatic chromosomes will help to support or refute the hypothesis that crossover landscapes in Arabidopsis hybrids are being shaped by the differential enrichment of MMR proteins between genomic regions. I predict that these experiments will reveal an enrichment of MSH2 around genes, and a relative depletion in the pericentromere and centromere, based on studies of mutation accumulation in wild type and *msh2-1* lines and the presence of a Tudor domain in Arabidopsis MSH6 (Belfield et al., 2018; Zhao et al., 2018). However, this prediction is in conflict with observations in mouse, where MSH2 appeared strongly enriched over the centromeres during meiotic prophase I (Kolas et al., 2005). At a finer scale, I would predict MSH2 ChIP-seq to reveal greater enrichment towards the 5'-end of gene bodies, in proximity to sites of H3K4me3 enrichment.

Furthermore, to address the functional relevance of the Tudor domain in the Arabidopsis MSH6 protein, it would be interesting to confirm this proteins ChIP-seq enrichment in regions of elevated H3K4me3. Indeed, site-directed mutagenesis could be performed on the MSH6 Tudor domain, and the functional relevance of this domain could then be assessed by creating transgenic plants in the *msh6* background. An alternative approach would be to assess the crossover landscape in wild type and *msh2-1* mutants in the context of a mutant background with altered H3K4me3.

Such meiotically expressed, epitope tagged lines could also be used for IP-mass-spectrometry experiments, in order to assess which components of the somatic MMR machinery are involved in meiotic MMR, and whether MSH2-heterodimers are physically associating with meiotic recombination proteins. These experiments, although challenging in Arabidopsis, could be performed in both Col-0 inbred and Col/Ler hybrid lines to identify protein associations dependent on interhomolog polymorphism.

One interesting line of inquiry will be to investigate if MMR plays an anti-crossover role during meiosis in more divergent plant hybrids. This will require the study of intra- or interspecies hybrids which display reduced fertility due to chromosome nondisjunction, and assessing whether MMR contributes to this phenotype. One interesting possibility will be to utilise a series of highly divergent African Arabidopsis accessions or to create interspecies hybrids within the Arabidopsis genus (Durvasula et al., 2017). For these experiments, an RNA-interference or CRISPR-Cas9 strategy could be adopted in order to knockout or knockdown *MSH2*, using an approach similar to that previously reported in tomato (Sarma et al., 2018). This will confirm whether the Arabidopsis MMR system is contributing to a species-hybridisation barrier, analogous to that previously observed for bacteria and yeast interspecies hybrids (Rayssiguier et al., 1989; Hunter et al., 1996).

As an alternative approach to validating the role of chromatin in shaping the distribution of MMR proteins across meiotic chromosomes, it will be informative to produce crossover maps in additional species. This model predicts that larger genomes with pronounced regional differences in interhomolog polymorphism would show redistribution patterns similar to those observed in the Arabidopsis *msh2-1* mutant. This would also address the current limitation of having to compare and contrast the role of MMR in the highly dissimilar budding yeast and Arabidopsis genomes, and would help to establish whether MMR has a general, conserved role in shaping the crossover landscape. To validate the model where meiotic MMR activity recruits components of the Class I DNA repair pathway, it would be informative to map crossovers in *mlh1* or *mlh3* mutant backgrounds in Arabidopsis, which this model predicts would redistribute in a pattern similar to *msh2-1*, albeit with crossovers at a lower overall level (Jackson et al., 2006; Dion et al., 2007; Cooper et al., 2018).

One advantage of the present study is that meiosis was not detrimentally disturbed as a result of interhomolog polymorphism in either wild type or *msh2-1*, based on fertility analysis, cytological analysis and total crossover numbers. Therefore, this provided a relatively ‘clean’ system in which to study the relationship between MMR, crossovers and polymorphism. Extending this line of reasoning, as gene synthesis, and ultimately chromosome synthesis, becomes more feasible it will be possible to construct organisms with identical proteomes, but divergent genome sequences. This would enable an investigation of the role of DNA sequence polymorphism, independently of *trans* modifiers of meiotic recombination. However, caveats of this approach, besides the technical difficulties, include the possibility of expression differences due to codon-usage effects, the potential loss of microRNA regulation, and the effects of changing sequence composition on features of the chromatin environment, such as nucleosome positioning.

Together with the data presented in this study, these and other experiments will continue to unravel the puzzle of MMR's role in regulating meiotic recombination. In addition, this work will have implications for our understanding of population genetic and broader evolutionary questions, through improving our understanding of the relationship between genetic diversity and meiotic recombination.

6.5 Acknowledgements

I would like to thank Dr Christophe Lambing and Dr Heïdi Serra for useful discussions. I would also like to thank Dr Timothy Cooper for sharing unpublished data and for constructive feedback on the proposed models.

Chapter Seven – Appendix

Appendix 7.1 – Genotyping markers used to assess the *msh2-1* T-DNA introgression for L*er*-0, CLC and Ct-1 backcross lines. Chromosome (Chr) number, marker type (either simple sequence length polymorphism (SSLP) or cleaved amplified polymorphic sequences (CAPS)), the restriction enzyme (Enz) used for CAPS analysis, SSLP product size or CAPS digestion product size in Col-0 and the corresponding accession, primer sequences, the marker's chromosome coordinates, and the closest gene (TAIR10 annotation) are provided.

Chr	Marker name	Type	Enz	Product size in Col-0	Product size in Ct-1	Forward primer	Reverse primer	Position (start bp) TAIR10	Closest gene
1	309	SSLP	-	271	224	CCAACAATTAGCCCGGATGA	CCAACAATTAGCCCGGATGA	309362	At1g01900
1	7294	SSLP	-	199	162	TTCAAAACTGGAGCGTCGTC	GGCCCATCTTGTGTGTTTTG	7294957	At1g20930
1	10655	SSLP	-	230	167	TTGTGGTCCCTGGCTAATCA	TTGTGGTCCCTGGCTAATCA	10655852	At1g30270
1	12356	SSLP	-	455	315	CTACGCCCGGTGTATTTGGA	CTACGCCCGGTGTATTTGGA	12356948	At3g30730
1	14122	SSLP	-	239	189	GCTAGCAGTCGAGTATTCTGTCGAG	GCTAGCAGTCGAGTATTCTGTCGAG	14122817	At1g37100
1	19077	SSLP	-	163	128	CCATCTTCTGTTTATTTGATTTCCA	CCATCTTCTGTTTATTTGATTTCCA	19076880	AT1G51450
1	23477	SSLP	-	183	129	TGCTTTTCCTTTTTAATCTTTTTCTCA	TGCTTTTCCTTTTTAATCTTTTTCTCA	23477122	At1g63295
1	27027	SSLP	-	365	278	ATCGGAATGCGGAAGACACT	ATCGGAATGCGGAAGACACT	27077150	At1g71930
1	30143	SSLP	-	135	104	CCAGCCACAGCTTCTTTCTGA	CCAGCCACAGCTTCTTTCTGA	30412519	At1g80950
2	132	SSLP	-	229	162	TCCAATGGGCCACAAATTAAC	TCCAATGGGCCACAAATTAAC	132648	At2g01250
2	2346	SSLP	-	347	261	GGCAAATTTGGTTGGCTCTC	GGCAAATTTGGTTGGCTCTC	2346993	At2g06020
2	4748	SSLP	-	291	237	TCGTCAAAACCGGAAAACCT	TCGTCAAAACCGGAAAACCT	4748219	At2g11810
2	6789	SSLP	-	112	82	GCGTTTTGTATCATCAAAGGTTCC	GCGTTTTGTATCATCAAAGGTTCC	6789815	At2g15560
2	9168	SSLP	-	218	152	ACCATGCCGCAAATGACATA	ACCATGCCGCAAATGACATA	9168713	At2g21410
2	15964	SSLP	-	171	120	TGCAGCACTGTGTTTTAATTTTAGTC	TGCAGCACTGTGTTTTAATTTTAGTC	15964454	At2g38120
2	17924	SSLP	-	585	329	CTGCTTCCACCAGAGAGTCC	CTGCTTCCACCAGAGAGTCC	17924327	At2g43110
2	19311	SSLP	-	140	101	TTTCTGCCAATGATTTAAAGTAACG	TTTCTGCCAATGATTTAAAGTAACG	19311521	At2g47000

3	259	SSLP	-	255	204	TGAGGCAATCCGGTTTTGAT	TGAGGCAATCCGGTTTTGAT	259868	At3g01710
3	1031	SSLP	-	419	345	ATGCCTTGGTTTCAATTTGG	ATGCCTTGGTTTCAATTTGG	1031481	At3g03980
3	1500	SSLP	-	371	295	TAGCCGCCGTCTACATAACC	TAGCCGCCGTCTACATAACC	1500250	At3g05270
3	2781	SSLP	-	223	185	ACAACTGGGCGACTCACCTT	ACAACTGGGCGACTCACCTT	2718687	At3g08940
3	5352	SSLP	-	533	311	GGTTCCAATCCACATATCTCTCC	GGTTCCAATCCACATATCTCTCC	5352423	At3g15820
3	9404	SSLP	-	384	297	AACGGTCCAGGTTCCTCCTC	AACGGTCCAGGTTCCTCCTC	9404279	At3g25760
3	10695	SSLP	-	161	122	GAGGGATGCAAGGAGGATCA	GAGGGATGCAAGGAGGATCA	10695968	At3g28540
3	11649	SSLP	-	228	188	TTTAGCCAAACATGCCCAAAT	TTTAGCCAAACATGCCCAAAT	11649496	At3g29770
3	12356	SSLP	-	455	315	CTACGCCCGGTGTATTTGGA	CTACGCCCGGTGTATTTGGA	12356948	At3g30730
3	15949	SSLP	-	465	382	CCACCCTCCAGGGAAGAAGT	CCACCCTCCAGGGAAGAAGT	15949551	At3g44250
3	17088	SSLP	-	560	360	GCTCTTGAGGTTTTAGGGTTGTT	GCTCTTGAGGTTTTAGGGTTGTT	17088210	At3g46430
3	19165	SSLP	-	284	234	TACGTCGCCCTCGAAGAAAT	TACGTCGCCCTCGAAGAAAT	19165521	At3g51660
3	21008	SSLP	-	211	172	CCGACGTTGTGTTTCTATTTCC	CCGACGTTGTGTTTCTATTTCC	21008135	At3g56710
3	22076	SSLP	-	231	171	TCGGAACTTACTTGACATATTCTACC	TCGGAACTTACTTGACATATTCTACC	22076576	At3g59765
3	23040	SSLP	-	228	180	TGCTACGACACGCAAACACA	TGCTACGACACGCAAACACA	23040094	At3g62260
4	230	SSLP	-	267	209	GCGTTCACCTTTAGCATTCCA	GCGTTCACCTTTAGCATTCCA	230388	At4g00520
4	4852	SSLP	-	146	108	TGGGCCAACGACTCTGTTTA	TGGGCCAACGACTCTGTTTA	4852373	At4g08028
4	6927	SSLP	-	168	129	TGAAAGGAGCATACCGTTGAGA	TGAAAGGAGCATACCGTTGAGA	6927139	At4g11385
4	10599	SSLP	-	559	357	TGGGTACATCTTAAAGGGTGGA	TGGGTACATCTTAAAGGGTGGA	10599322	At4g19430
4	18526	SSLP	-	478	319	GACGAACAAGGCAACCCATT	GACGAACAAGGCAACCCATT	18526361	At4g39950
5	53	SSLP	-	133	93	TCTGCATGGGAAATCTCTGG	TCTGCATGGGAAATCTCTGG	53020	At5g01150
5	7064	SSLP	-	267	220	ACTGGCCTCGCCTTTCACTA	ACTGGCCTCGCCTTTCACTA	7064379	At5g20840
5	10406	SSLP	-	350	272	TGTATAATTAGAGCCGTTCGTCGT	TGTATAATTAGAGCCGTTCGTCGT	10406321	At5g28468
5	16428	SSLP	-	229	183	TGTTGCCATGTTGATTTGATTG	TGTTGCCATGTTGATTTGATTG	16428466	At5g41030
5	19994	SSLP	-	169	109	TCTAAACCGAACTAAACCGTGAA	TCTAAACCGAACTAAACCGTGAA	19994907	At5g49320
5	23287	SSLP	-	204	151	GAGATGTTGAGAAGCAGAGGAAA	GAGATGTTGAGAAGCAGAGGAAA	23287613	At5g57500
5	26907	SSLP	-	270	200	TGTGGATCTTTATGACGTGTGC	TGTGGATCTTTATGACGTGTGC	26907352	At5g67420

Chr	Marker name	Type	Enz	Product size in Col-0	Product size in CLC	Forward primer	Reverse primer	Position (start bp) TAIR10	Closest gene
1	1505	CAPS	HindIII	126+227	353	CCAAGCTCTGTGAATGTTGGAGC	GATTGTGGACCAATGGCTGC	1505941	At1g05200
1	3298	CAPS	HindIII	229+100	329	AATACAACACAGCTATCGGTGC	TTAATGTCCAGCGGCTGATGC	3298485	At1g10095
1	5292	CAPS	HindIII	95+95	190	TGTGCTTGGTTTTGGGTGTA	AGTTGCATTGCACAAAAGAGTT	5292435	At1g15380
1	6108	CAPS	HindIII	77+154	231	CAGTTCACGGGTCCAATACC	CCGGCATAAAACCAAAAAGA	6108789	At1g17750
1	7294	SSLP	-	199	162	TTCAAAACTGGAGCGTCGTC	GGCCCATCTTGTGTGTTTTG	7294957	At1g20930
1	10443	SSLP	-	245	141	TTCTCGATGGGATGATAGGTG	CCAATCAATGACCAACAAAAAG	10443987	At1g29830
1	10655	SSLP	-	230	167	TTGTGGTCCCTGGCTAATCA	TTGTGGTCCCTGGCTAATCA	10655852	At1g30270
1	11184	CAPS	HindIII	31+214	245	TTGAGGAATTGCTTCGATCC	GACATTGCTTCCACCAACCT	11184029	At1g31280
1	13278	CAPS	HindIII	23+151	174	TCGCATCTGAGAATTGCTTG	GATTGTTTCAGCAACCACCA	13278246	At1g35780
1	14122	SSLP	-	239	189	GCTAGCAGTCGAGTATTCTGTCGAG	GCTAGCAGTCGAGTATTCTGTCGAG	14122817	At1g37100
1	15621	CAPS	HindIII	165+165	330	GAAACTTCATGTTGCTAAGAGAGC	GGAGTTCAAATGTGAGACTTGCC	15621126	At1g41840
1	16908	SSLP	-	110	82	GCACAGAAAGACAAACCCAAAG	CGACCAGCAAGGTTGTTCTTAG	16907783	At1g44770
1	18587	CAPS	HindIII	82+83	165	TGATTGCTCGTAGCATGTGA	AATCTCAAAGACGACGCAAA	18587671	At1g50190
1	19540	SSLP	-	221	286	GTTCCCCGATTCATGTGAGA	CAAAAAGGGAAAAGCCCACT	19539729	At1g52440
1	20074	CAPS	HindIII	65+85	150	CTGCCTACACCGTCATCAAA	TCCTTCTCGCCATCTCAGTT	20074913	At1g53780
1	21975	CAPS	HindIII	119+39	158	TTGCTTTTGAATTTATGAGTGGAA	GAATATTTGCCAAGCCATCG	21975829	At1g59750
1	24743	SSLP	-	234	160	GAGGCACCGAAAATGGATTA	CCAATCGGATTATAGTGTGAACTTT	24743355	At1g66345
1	25323	CAPS	HindIII	73+79	152	AATGCATCCGGTTTACAAGC	ACGCTGCAGAGCTAAGTTCC	25323140	At1g67560
1	26352	SSLP	-	231	131	CATAAGAGCCCCGATACTACTCA	CAAGGAGATGTTGGGCTTTG	26357768	At1g69980
1	30413	SSLP	-	135	104	CCAGCCACAGCTTCTTTCTGA	TTGATTGAATAATGGTTCTTGTGATGA	30412519	At1g80950
2	132	SSLP	-	229	162	TCCAATGGGCCACAAATTAAC	TTTGTGCTTTGATTACTGCAAGTG	132648	At2g01250
2	1027	CAPS	HindIII	193+194	387	TGGAGTCCCTGGTTGTGTGC	ACATGATAAGCGTCGATAATAAGACG	1027654	At2g03370
2	2346	SSLP	-	347	261	GGCAAATTTGGTTGGCTCTC	TGTTTTGTGCTATTTGTGTCAACC	2346993	At2g06020
2	3098	CAPS	HindIII	198+117	315	CCTAGAAACCACCATCAGAGAGC	TTGGCTAGGAAAGCCTTCTTGC	3098691	At2g07450
2	4302	SSLP	-	283	163	CCAACGTCACCTCCTCCTTA	AACGGCTATGACTATCCAATTAAGA	4302041	At2g10921
2	6789	SSLP	-	112	82	GCGTTTTGTATCATCAAAGGTTCC	CGCAATTTCTCGAACTTCCTTT	6789815	At2g15560

2	10268	CAPS	BamHI	160+40	187	TCATTTGAAGCCAGTTGGAA	TTTCCATCTTGGTTGGGTTC	10268940	At2g24160
2	11321	CAPS	EcoRI	81+79	160	TGAAACTCCAAACGCATCAG	TTACCAGTGCAAATGCGAAA	11321269	At2g26610
2	12121	SSLP	-	229	157	CAACGAAGAGAGGGAAACAAA	CGTTTCCAAATGACAAATGGT	12121783	At2g28355
2	13427	CAPS	PstI	140+65	205	GTGGAGATTATTCCGCCAGA	CAAGCAAAACAGCCGTTACA	13427779	At2g31530
2	14176	SSLP	-	205	146	TTGAGAGATTTTGCAATAGTAAGCA	CGAATCAATCTTATCAACTCTTCTTG	14176271	At2g33470
2	15642	CAPS	HindIII	21+147	168	TCGATTTGGTGGTGAATTTG	GATGTGGAGGGAGAAGTGGA	15642866	At2g37250
2	16339	SSLP	-	202	142	CACGAGCAATCCTTGTTTCA	GGGAAAAAGAAAGACCCACA	16339573	At2g39175
2	17165	CAPS	BamHI	57+130	187	TCATCAACGTTGCTCATAAACC	CGTCTTGACCGGTGAGTTCT	17165057	At2g41180
2	18087	CAPS	HindIII	207+104	311	ATGCAGTTCTGCGAGAACACG	TCCAGGGTTAATAGTTGACTTGCG	18087124	At2g43600
2	19048	SSLP	-	306	230	GACCCATATCGTAGGCCACT	TTTTACCAGCCTCCATCGAC	19048891	At2g46410
3	1031	SSLP	-	419	345	ATGCCTTGGTTTCAATTTGG	TACCCGCTCCTTGACAGTTT	1031481	At3g03980
3	1746	CAPS	HindIII	23+146	169	TGAATGCACAGTCAGAGCTAAA	TTTCTAGAGATTACCTCCCTTTTTGA	1746235	At3g05850
3	2718	SSLP	-	223	185	ACAACTGGGCGACTCACCTT	CGTAAACACAAACTGCGAGGT	2718687	At3g08940
3	3621	CAPS	HindIII	121+31	152	ACGTACCACCATCCCAATGT	TTTTTGTTGTTACCCCCAATG	3621888	At3g11500
3	4126	SSLP	-	310	169	GGAATAATGGATTCCTCTCTCG	TGTTTGAATGTTGACAATGAGC	4126508	At3g12930
3	4715	CAPS	HindIII	176+22	198	TGGCTTGTAGGTGTTGTTGAA	CAAAGCAGCCATTGATGATG	4715997	At3g14205
3	5560	CAPS	HindIII	132+71	203	CTTGATTTCCTGCCATAGCTGC	TACAGAGCAATCTGCTGTATCTGC	5560232	At3g16380
3	7638	CAPS	HindIII	105+81	186	ACGGCGAGCTAGAAACTGTC	GTTCGAGGTTACCGAGATGG	7638911	At3g21690
3	8935	CAPS	HindIII	114+36	150	CCAAAACCAACCACTGCTTT	TGAGAAGTCTGGTGAAAGTGGA	8935217	At3g24515
3	9404	SSLP	-	384	297	AACGGTCCAGGTTCCTCCTC	TTGGTTTTAAGGCTCTGGAATCA	9404279	At3g25760
3	10695	SSLP	-	161	122	GAGGGATGCAAGGAGGATCA	TTCATCACATCAACGCTCCAA	10695968	At3g28540
3	11649	SSLP	-	228	188	TTTAGCCAAACATGCCCAAAT	CCAAGCGCCAAAACTACCTC	11649496	At3g29770
3	12356	SSLP	-	455	315	CTACGCCCGGTGTATTTGGA	GCTTGTGAGGCTATGTGGCTTA	12356948	At3g30730
3	14278	CAPS	HindIII	125+87	212	ATATTTGCAGCACCACATGGACC	TTCAAACCAGCTATGCATACCG	14278309	At3g42100
3	15949	SSLP	-	465	382	CCACCCTCCAGGGAAGAAGT	GGCAGCGACTGGCTTGTTTA	15949551	At3g44250
3	16679	CAPS	HindIII	122+69	191	ATCCATTAGCAAGGCGATGT	CCGTGAGTTTGGGAATCAAT	16679705	At3g45470
3	17233	CAPS	HindIII	159+31	190	CACAGAACCCAAATCCGAGT	GATTCGCGTCTTGTCTCAGG	17233598	At3g46800
3	18459	CAPS	HindIII	114+42	156	TCAGAGACAGCAGAGGTGTGA	CGAAGATTTGCGAGAGAACA	18459071	At3g49765

3	19165	SSLP	-	284	234	TACGTCGCCCTCGAAGAAAT	GCGCTACATACGCACCACAT	19165521	At3g51660
3	19604	CAPS	HindIII	124+52	176	TTGTCGCTGAAGTTGGTTTG	AACCACGACGGTTGGATAAA	19604527	At3g52880
3	20226	CAPS	HindIII	148+61	209	CCAACTTCCTTCTCTTCCCTGC	TCGGCTGGTGAGAACCATATGC	20226178	At3g54650
3	21008	SSLP	-	211	172	CCGACGTTGTGTTTCTATTTCC	TGAGGGAACAAGGACCTAACCA	21008135	At3g56710
3	22117	CAPS	HindIII	207+56	263	TCCACGTCGTCTTGATCTCTCG	TTATTCATTGCATTGAATAAGACGATGC	22117660	At3g59870
3	23335	CAPS	HindIII	163+86	249	TTCTCGTTTCCCGCTTCAATCG	AGTGAGATCTGTTACTTAGACTTCCG	23335087	At3g63170
4	230	SSLP	-	267	209	GCGTTCACCTTTAGCATTCCA	GCAGCTACACTCATGCCCTCT	230388	At4g00520
4	2611	SSLP	-	243	201	CTGGAGGAAAGGTTGGTGAA	TGAGCCTCCTTTCTGATTGA	2611319	At4g05091
4	3477	CAPS	HindIII	114+208	322	CCACGAACTTCACCAACCTAGC	GAATTAGATGCAAGTGTGGCTGC	3477673	At4g06557
4	4852	SSLP	-	146	108	TGGGCCAACGACTCTGTTTA	TGGGCCAACGACTCTGTTTA	4852373	At4g08028
4	7157	SSLP	-	334	262	ACATTAGCGGAGGCCACTT	ATGGGCAAAAGCTTCCAGTA	7157357	At4g11911
4	8198	CAPS	HindIII	81+163	244	ACAAAACCGAACCCCAAAAG	GCTCTGAACAATGCAGGATG	8198181	At4g14220
4	9210	CAPS	HindIII	86+134	220	ATGAGGAGACCGGGGTAAGT	AGTTGGGAACCCTGTTCCTT	9210030	At4g16280
4	10721	SSLP	-	323	188	ACAATTTTTAGTCTGTCTAGCGTGA	CGAAATGCAGTTCACATCGT	10721530	At4g19710
4	11124	CAPS	HindIII	116+41	157	TCGATAACCACTTAATTGTTGAGA	CCAAATGTCTCATCTCGTCGT	11124218	At4g20740
4	16133	CAPS	HindIII	208+100	308	TTGGCTAATACCGGGTTAGTGC	CGCGTGGGAACTAACTGACAC	16133673	At4g33540
5	1414	CAPS	HindIII	113+86	199	GAGATAGAGAGAGAAAAGGACGGTAA	AAGAAAAGAAGGTAATCACACAACG	1414820	At5g04860
5	2628	CAPS	HindIII	111+43	154	TGCCATAATGCAAGCAAAGT	ATGAATCCTGGCCGTTGATA	2628593	At5g08170
5	3750	SSLP	-	137	97	ATGGTGGACCTGGGGGTAAC	GCATGTAGGAAACACAAATCCTGA	3750331	At5g11660
5	7064	SSLP	-	267	220	ACTGGCCTCGCCTTTCACTA	AATCACAACTGTGCCCTCGTT	7064379	At5g20840
5	8569	CAPS	HindIII	61+156	217	ATTTGCCACAGATCCCAAAA	GTCGCTATTTCGTGGAAACG	8569949	At5g24910
5	9437	CAPS	HindIII	128+62	190	TCGTTTCACAACTTCTTCTTCG	CGGATCGTGTGGAAGAGACT	9437511	At5g26820
5	16428	SSLP	-	229	183	TGTTGCCATGTTGATTTGATTG	TGTTGCCATGTTGATTTGATTG	16428466	At5g41030
5	19994	SSLP	-	169	109	TCTAAACCGAACTAAACCGTGAA	CAAACCAAAACCTACTTTTTCCAA	19994907	At5g49320
5	20685	CAPS	EcoRI	61+149	210	CCGAGTTAAGCCCAATTTGA	TCAATTCGAACCGAAAGCAT	20685339	At5g50830
5	21349	CAPS	EcoRI	109+41	150	TGTATTTTGGATTTTGGTTCCAG	TCGTCCAGCCTTTTAGTTGC	21349815	At5g52630
5	22402	CAPS	HindIII	110+56	166	GGGTCCCACACTCACCTCT	GGGTTTAAAATGGGTTTCTCTTT	22402070	At5g55230
5	24192	CAPS	HindIII	102+122	224	TCAAATGGGATCAAAAACAACA	GGATTCGAGTTCCACGAGAA	24192726	At5g60070

5	25212	SSLP	-	159	123	GCGGTGGCAGTAGGTTAAAA	TCTAATACCGGCAATAAAACTTGA	25212874	At5g62780
5	26907	SSLP	-	270	200	TGTGGATCTTTATGACGTGTGC	ACCATCTACTTCCATTCAAATAACG	26907352	At5g67420
Chr	**Marker name**	**Type**	**Enz**	**Product size in Col-0**	**Product size in *Ler*-0**	**Forward primer**	**Reverse primer**	**Position (start bp) TAIR10**	**Closest gene**
1	309	SSLP	-	271	224	CCAACAATTAGCCCGGATGA	CCAACAATTAGCCCGGATGA	309362	At1g01900
1	1426	SSLP	-	330	247	CGTGTTGTTGTTGTTGTGTGTGG	CGTGTTGTTGTTGTTGTGTGTGG	1427801	At1g05000
1	3025	SSLP	-	271	227	GGTGCCACCATGTAGATTCGG	GGTGCCACCATGTAGATTCGG	3031801	At1g09390
1	3830	SSLP	-	130	98	CAACAATGGTGATATTTGTTTTGC	CAACAATGGTGATATTTGTTTTGC	3830600	At1g11370
1	9200	SSLP	-	214	165	GGACAATTAATCATGCTCAATTGCTC	GGACAATTAATCATGCTCAATTGCTC	9203801	At1g26630
1	10655	SSLP	-	230	167	TTGTGGTCCCTGGCTAATCA	TTGTGGTCCCTGGCTAATCA	10655852	At1g30270
1	12356	SSLP	-	455	315	CTACGCCCGGTGTATTTGGA	CTACGCCCGGTGTATTTGGA	12356948	At3g30730
1	14122	SSLP	-	239	189	GCTAGCAGTCGAGTATTCTGTCGAG	GCTAGCAGTCGAGTATTCTGTCGAG	14122817	At1g37100
1	17135	SSLP	-	482	398	CCTATGTTTCGGCATTCAGG	CCTATGTTTCGGCATTCAGG	17135019	At1g45211
1	19077	SSLP	-	163	128	CCATCTTCTGTTTATTTGATTTCCA	CCATCTTCTGTTTATTTGATTTCCA	19076880	At1g51450
1	21236	SSLP	-	476	340	CAATGAGCCCTCTACGCTCT	CAATGAGCCCTCTACGCTCT	21236506	At1g56650
1	27027	SSLP	-	365	278	ATCGGAATGCGGAAGACACT	ATCGGAATGCGGAAGACACT	27077150	At1g71930
1	30143	SSLP	-	135	104	CCAGCCACAGCTTCTTTCTGA	CCAGCCACAGCTTCTTTCTGA	30412519	At1g80950
2	132	SSLP	-	229	162	TCCAATGGGCCACAAATTAAC	TCCAATGGGCCACAAATTAAC	132648	At2g01250
2	2346	SSLP	-	347	261	GGCAAATTTGGTTGGCTCTC	GGCAAATTTGGTTGGCTCTC	2346993	At2g06020
2	3689	SSLP	-	126	75	AATCCGCTCTCACTGAGCAT	AATCCGCTCTCACTGAGCAT	3689934	At2g09830
2	4302	SSLP	-	283	163	CCAACGTCACCTCCTCCTTA	AACGGCTATGACTATCCAATTAAGA	4302041	At2g10921
2	6789	SSLP	-	112	82	GCGTTTTGTATCATCAAAGGTTCC	GCGTTTTGTATCATCAAAGGTTCC	6789815	At2g15560
2	8300	SSLP	-	243	195	TTACGGTGTTTTGGATCCGTTAGC	TTACGGTGTTTTGGATCCGTTAGC	8304401	At2g19146
2	9168	SSLP	-	218	152	ACCATGCCGCAAATGACATA	ACCATGCCGCAAATGACATA	9168713	At2g21410
2	11443	SSLP	-	200	141	GGTTCCGTCAACTTCGAAAA	GGTTCCGTCAACTTCGAAAA	11443153	At2g26830
2	13036	SSLP	-	231	231	ACATACATATGCCCAAAATTGTTATC	ACATACATATGCCCAAAATTGTTATC	13036324	At2g30600
2	14714	SSLP	-	125	89	CAATTAAAGAGGTTTCAGTTTTCCAG	CAATTAAAGAGGTTTCAGTTTTCCAG	14714870	At2g34880
2	15964	SSLP	-	171	120	TGCAGCACTGTGTTTTAATTTTAGTC	TGCAGCACTGTGTTTTAATTTTAGTC	15964454	At2g38120

2	16339	SSLP	-	202	142	CACGAGCAATCCTTGTTTCA	GGGAAAAAGAAAGACCCACA	16339573	At2g39175
2	18110	SSLP	-	253	197	ATCTTGCATTTCCTCCTTGGTGTC	ATCTTGCATTTCCTCCTTGGTGTC	18115501	At2g43700
2	19554	SSLP	-	175	134	CACACGAATATTGATTGTCTAAGGA	CACACGAATATTGATTGTCTAAGGA	19554916	At2g47710
3	1031	SSLP	-	419	345	ATGCCTTGGTTTCAATTTGG	ATGCCTTGGTTTCAATTTGG	1031481	At3g03980
3	1500	SSLP	-	371	295	TAGCCGCCGTCTACATAACC	TAGCCGCCGTCTACATAACC	1500250	At3g05270
3	2718	SSLP	-	223	185	ACAACTGGGCGACTCACCTT	CGTAAACACAAACTGCGAGGT	2718687	At3g08940
3	3520	SSLP	-	229	173	CTCGGCTTCGCATCTAGTTC	CTCGGCTTCGCATCTAGTTC	3520343	At3g11240
3	4126	SSLP	-	310	169	GGAATAATGGATTCCTCTCTCG	TGTTTGAATGTTGACAATGAGC	4126508	At3g12930
3	4868	SSLP	-	295	157	CAAACTTGAGCAGGATCTGGTCG	CAAACTTGAGCAGGATCTGGTCG	4873201	At3g14520
3	7120	SSLP	-	242	188	GAGATGTGTGTTATGCCGGTAGC	GAGATGTGTGTTATGCCGGTAGC	7123101	At3g20430
3	7595	SSLP	-	177	132	AACGAAAAAGGGGGAATATGAA	AACGAAAAAGGGGGAATATGAA	8495301	At3g23633
3	8495	SSLP	-	177	132	AACGAAAAAGGGGGAATATGAA	AACGAAAAAGGGGGAATATGAA	8495131	At3g23633
3	9404	SSLP	-	384	297	AACGGTCCAGGTTCCTCCTC	AACGGTCCAGGTTCCTCCTC	9404279	At3g25760
3	10695	SSLP	-	161	122	GAGGGATGCAAGGAGGATCA	GAGGGATGCAAGGAGGATCA	10695968	At3g28540
3	11649	SSLP	-	228	188	TTTAGCCAAACATGCCCAAAT	TTTAGCCAAACATGCCCAAAT	11649496	At3g29770
3	12356	SSLP	-	455	315	CTACGCCCGGTGTATTTGGA	CTACGCCCGGTGTATTTGGA	12356948	At3g30730
3	15949	SSLP	-	465	382	CCACCCTCCAGGGAAGAAGT	CCACCCTCCAGGGAAGAAGT	15949551	At3g44250
3	17088	SSLP	-	560	360	GCTCTTGAGGTTTTAGGGTTGTT	GCTCTTGAGGTTTTAGGGTTGTT	17088210	At3g46430
3	19165	SSLP	-	284	234	TACGTCGCCCTCGAAGAAAT	TACGTCGCCCTCGAAGAAAT	19165521	At3g51660
3	21008	SSLP	-	211	172	CCGACGTTGTGTTTCTATTTCC	CCGACGTTGTGTTTCTATTTCC	21008135	At3g56710
3	22076	SSLP	-	231	171	TCGGAACTTACTTGACATATTCTACC	TCGGAACTTACTTGACATATTCTACC	22076576	At3g59765
3	23040	SSLP	-	228	180	TGCTACGACACGCAAACACA	TGCTACGACACGCAAACACA	23040094	At3g62260
4	230	SSLP	-	267	209	GCGTTCACCTTTAGCATTCCA	GCGTTCACCTTTAGCATTCCA	230388	At4g00520
4	4852	SSLP	-	146	108	TGGGCCAACGACTCTGTTTA	TGGGCCAACGACTCTGTTTA	4852373	At4g08028
4	9652	SSLP	-	234	172	GTTGCCCACTTGTGTGGTCT	GTTGCCCACTTGTGTGGTCT	9652287	At4g17200
4	10599	SSLP	-	559	357	TGGGTACATCTTAAAGGGTGGA	TGGGTACATCTTAAAGGGTGGA	10599322	At4g19430
4	11733	SSLP	-	193	147	CGTGTGCTTAGCCAGAAACA	CGTGTGCTTAGCCAGAAACA	11732797	At4g22150
4	12303	SSLP	-	189	154	TGGAGTTAAAAGTCAAAGAATTGAG	TGGAGTTAAAAGTCAAAGAATTGAG	12302837	At4g23570

4	14558	SSLP	-	168	136	AAATCAAAACCCCATGAAAGG	AAATCAAAACCCCATGAAAGG	14558575	At4g29730
4	16046	SSLP	-	153	115	GGCGCTAATGTACACTCTTCG	GGCGCTAATGTACACTCTTCG	16046992	At4g33280
4	18526	SSLP	-	478	319	GACGAACAAGGCAACCCATT	GACGAACAAGGCAACCCATT	18526361	At4g39950
5	53	SSLP	-	133	93	TCTGCATGGGAAATCTCTGG	TCTGCATGGGAAATCTCTGG	53020	At5g01150
5	332	SSLP	-	258	191	CCACTTCACACATGGCTACTG	CCACTTCACACATGGCTACTG	331691	At5g01849
5	1530	SSLP	-	184	142	ATCTTAAGATTTCTTTAGTAAGAAGCGC	ATCTTAAGATTTCTTTAGTAAGAAGCGC	1538601	At5g05180
5	7064	SSLP	-	267	220	ACTGGCCTCGCCTTTCACTA	ACTGGCCTCGCCTTTCACTA	7064379	At5g20840
5	9005	SSLP	-	154	124	AGAAGTCCCAATACATGCATGAAGAG	AGAAGTCCCAATACATGCATGAAGAG	9007701	At5g25840
5	10406	SSLP	-	350	272	TGTATAATTAGAGCCGTTCGTCGT	TGTATAATTAGAGCCGTTCGTCGT	10406321	At5g28468
5	16428	SSLP	-	229	183	TGTTGCCATGTTGATTTGATTG	TGTTGCCATGTTGATTTGATTG	16428466	At5g41030
5	19994	SSLP	-	169	109	TCTAAACCGAACTAAACCGTGAA	TCTAAACCGAACTAAACCGTGAA	19994907	At5g49320
5	23287	SSLP	-	204	151	GAGATGTTGAGAAGCAGAGGAAA	GAGATGTTGAGAAGCAGAGGAAA	23287613	At5g57500
5	26907	SSLP	-	270	200	TGTGGATCTTTATGACGTGTGC	TGTGGATCTTTATGACGTGTGC	26907352	At5g67420

Appendix 7.2 – List of oligonucleotides used for T-DNA genotyping, construct assembly and qPCR analyses.

Primer name	Sequence 5′ – 3′	Purpose
SALK_0020708R	CCTCCCATGTTAGGCCCTGTT	SALK *MSH2* WT genotyping
SALK_0020708L	AGCGCAATTTGGGCATGTCT	
SALK_136296_RP	TGGAATGGATCAATGAGTTCC	SALK *MSH4* WT genotyping
SALK_136296_LP	CGGCTTCACTGCATCTATCTC	
LBb1.3_SALK	ATTTTGCCGATTTCGGAAC	SALK *msh2-1* & *msh4-1* T-DNA border primer
LMN_F	TAAGCACTGCAGATCCAACTCGAGGGCG	Primers for amplifying 'Linker-Myc×3-NOS' region from a previously reported vector (Choi et al., 2018)
LMN_R	TGCTTAGGATCCCCCCGATCTAGTAACATAG ATGAC	
HA_tag_A	GATCCATCCAACTCGAGGGCGCGCCTGTAC AGACGTCTC	Oligonucleotides for annealing 'Linker-HA×3-NOS' with BamHI & KpnI sites
HA_tag_B	ATCGATACCGTCGAGACGTCTGTACAGGCG CGCCCTCGAGTTGGATG	
HA_tag_C	GACGGTATCGATTCAAAGCTATGGAGTACCC TTACGATGTGCCTGACTAC	
HA_tag_D	ATATGGATAAGCGTAGTCAGGCACATCGTAA GGGTACTCCATAGCTTTGA	
HA_tag_E	GCTTATCCATATGACGTTCCAGATTACGCTT ACCCATATGATGTT	
HA_tag_F	CTCTTCACCCGGGAGCGTAATCAGGAACAT CATATGGGTAAGCGTAATCTGGAACGTC	
HA_tag_G	CCTGATTACGCTCCCGGGTGAAGAGGTAC	
MSH4_Ct_1F	GGTCGACGGTATCGATAAGCTTGATGGACT ATGTGTAGTGTTTACGC	Amplifying the *MSH4* promoter and gene body in three ~2.4 kb amplicons for Gibson assembly
MSH4_Ct_1R	GACACCATTTGATTGTAGC	
MSH4_Ct_2F	CAGCATTATCTCTCATTGG	
MSH4_Ct_2R	CAACTAAATATGATTAAGCACC	
MSH4_Ct_3F	CAGATGCCTCGTTTCC	
MSH4_Ct_3R	CGAGTTGGATCTGCAGGAATTCGATGAGTCT TTCTTCAGTGAAGC	
MSH4_prom_R	CGAGTTGGATCTGCAGGAATTCGATTTTCGC TCCACAGATCAG	Amplifying *MSH4* promoter for Gibson assembly
MSH4_GB_Myc_F 1	GAAGAGGACTTGAATTCGGTACCCCATGGA AGACGACGGAGGAG	Amplifying the *MSH4* gene body in two ~2.7 kb amplicons for Gibson assembly
MSH4_GB_Myc_R 1	GCAAGTTGAATTCCTCACG	
MSH4_GB_Myc_F 2	CAATCTTGTTAACACTGGC	
MSH4_GB_Myc_R 2	TCGGGGAAATTCGAGCTCGGTACCCTTAGA GTCTTTCTTCAGTGAAGC	
MSH4_GB_HA_F1	ATATGATGTTCCTGATTACGCTCCCATG GAAGACGACGGAGGAG	Amplifying the *MSH4* gene body in two ~2.7 kb amplicons for Gibson assembly
MSH4_GB_HA_R1	GCAAGTTGAATTCCTCACG	
MSH4_GB_HA_F2	CAATCTTGTTAACACTGGC	
MSH4_GB_HA_R2	CGATGAGCTCGGTACCTCTTCACCCGAGTC TTTCTTCAGTGAAGC	
T1_HA-MSH4_F	TCCCTTTCTTGGTTAAGGTAGGC	Genotyping HA-MSH4 transgene
HEI10_OE_5′_F	CAGGAAACAGCTATGACCATGATTACG	Genotyping HEI10 overexpressor
HEI10_OE_5′_R	CCTTAACAATGAGATGCAAGTCTACG	

RAC1_qPCR_F	GCCCTGTCGATGAAAGTGCT	RAC1 exon 1 qPCR primers
RAC1_qPCR_R	ATCGGGTTCCTCAGGGATTG	
Ta2_JP1725_F	AAACGATGCGTTGGGATAGGTC	Ta2 transposon qPCR primers
Ta2_JP1726_R	ATACTCTCCACTTCCCGTTTTTCTTTTTA	
HP3_ChIP_F2	TGTTCGTTAATTCTTGAGCCTCA	3a hotspot qPCR primers (set 1)
HP3_ChIP_R2	TGGCTGAATCGTTGACTAGGA	
HP4_ChIP_F1	TTGGGGTTTGAGTTTGCCAC	3a hotspot qPCR primers (set 2)
HP4_ChIP_R1	TGTCAGTCTCCAACCTCGTT	
HP4_ChIP_F2	ACGGCCTTCTTTAGTCTTCCA	3a hotspot qPCR primers (set 3)
HP4_ChIP_R2	GTAGCTTCCGTTGTTCAGCA	
HP5_ChIP_F2	ACTAGGCTGCGTCGTGTAAA	3b hotspot qPCR primers (set 1)
HP5_ChIP_R2	GGAGTTTTGTTCCGTGGCTC	

Appendix 7.3 – Flow cytometry fluorescent pollen count data for *I1b*, *I2f* and *I5a* intervals in Col-0 inbred wild type and *msh2-1* mutant backgrounds. Genetic distance in centimorgans (cM) is calculated as $cM = 100\times(N_Y/(N_Y+N_{R+Y}))$, where N_Y is the number of yellow alone pollen, and N_{Y+R} is the number of red and yellow pollen. Mann-Whitney *U* tests were performed to test for significant differences between genotypes.

Individual	Total	Red alone	Red + Yellow	Non-colour	Yellow alone	cM	cM/Mb	*P*-value
I1b Col-0 WT	24,624	1,104	10,015	12,651	854	7.86	4.25	
I1b Col-0 WT	21,161	938	8,034	11,416	773	8.78	4.74	
I1b Col-0 WT	24,760	1,052	10,053	12,742	913	8.33	4.50	
I1b Col-0 WT	21,222	901	8,020	11,565	736	8.41	4.54	
I1b Col-0 WT	23,913	1,087	10,105	11,771	950	8.59	4.65	
Total	**115,680**	**5,082**	**46,227**	**60,145**	**4,226**	**8.39**	**4.54**	**-**
I1b Col-0 *msh2*	26,220	1,187	10,021	14,036	976	8.88	4.80	
I1b Col-0 *msh2*	20,792	908	8,019	11,094	771	8.77	4.74	
I1b Col-0 *msh2*	8,610	365	3,188	4,746	311	8.89	4.80	
Total	**55,622**	**2,460**	**21,228**	**29,876**	**2,058**	**8.84**	**4.78**	**0.071**
I2f Col-0 WT	23,559	1,080	9,245	12,456	778	7.76	11.59	
I2f Col-0 WT	16,764	882	5,303	9,793	786	12.91	19.27	
I2f Col-0 WT	26,522	1,417	9,137	14,877	1,091	10.67	15.92	
I2f Col-0 WT	25,129	1,167	9,223	13,840	899	8.88	13.26	
I2f Col-0 WT	25,431	1,184	9,153	14,005	1,089	10.63	15.87	
Total	**117,405**	**5,730**	**42,061**	**64,971**	**4,643**	**10.17**	**15.18**	**-**
I2f Col-0 *msh2*	25,366	1,138	9,363	14,189	676	6.73	10.05	
I2f Col-0 *msh2*	35,771	806	7,095	27,457	413	5.50	8.21	
I2f Col-0 *msh2*	26,079	1,230	9,221	14,599	1,029	10.04	14.98	
I2f Col-0 *msh2*	25,989	1,025	8,780	14,522	1,662	15.92	23.76	
I2f Col-0 *msh2*	17,337	831	6,235	9,639	632	9.20	13.74	
Total	**130,542**	**5,030**	**40,694**	**80,406**	**4,412**	**9.48**	**14.15**	**0.548**
I5a Col-0 WT	19,198	2,015	6,459	9,038	1,686	20.70	4.21	
I5a Col-0 WT	23,430	2,528	8,039	10,621	2,242	21.81	4.43	
I5a Col-0 WT	23,451	2,537	8,025	10,588	2,301	22.28	4.53	
I5a Col-0 WT	24,199	2,542	8,029	11,318	2,310	22.34	4.54	
I5a Col-0 WT	21,565	2,077	7,009	10,625	1,854	20.92	4.25	
I5a Col-0 WT	20,886	2,172	6,998	9,797	1,919	21.52	4.37	
I5a Col-0 WT	23,937	2,509	8,018	11,278	2,132	21.00	4.27	
Total	**156,666**	**16,380**	**52,577**	**73,265**	**14,444**	**21.51**	**4.37**	**-**
I5a Col-0 *msh2*	24,517	2,557	8,013	11,565	2,382	22.91	4.66	
I5a Col-0 *msh2*	23,892	2,086	6,505	13,485	1,816	21.82	4.44	
I5a Col-0 *msh2*	30,032	3,032	10,029	14,192	2,779	21.70	4.41	
I5a Col-0 *msh2*	14,988	1,574	5,003	7,125	1,286	20.45	4.16	
I5a Col-0 *msh2*	15,700	1,792	5,007	7,492	1,409	21.96	4.46	
I5a Col-0 *msh2*	13,831	1,303	4,000	7,398	1,130	22.03	4.48	
Total	**122,960**	**12,344**	**38,557**	**61,257**	**10,802**	**21.81**	**4.43**	**0.534**

Appendix 7.4 – *I1b* flow cytometry fluorescent pollen count data for the wild type and *msh2-1* mutant in hybrid backgrounds. Genetic distance in centimorgans (cM) is calculated as cM = $100\times(N_Y/(N_Y+N_{R+Y}))$, where N_Y is the number of yellow alone pollen, and N_{Y+R} is the number of red and yellow pollen. Mann-Whitney *U* tests were performed to test for significant differences between genotypes.

Individual	Total	Red alone	Red + Yellow	Non-colour	Yellow alone	cM	cM/Mb	*P*-value
Col/Ler F_1 WT	17,212	719	4,455	8,595	235	5.01	2.71	
Col/Ler F_1 WT	13,382	473	4,064	6,346	254	5.88	3.18	
Col/Ler F_1 WT	24,994	960	8,147	13,308	441	5.14	2.78	
Col/Ler F_1 WT	24,563	990	7,969	13,422	481	5.69	3.08	
Col/Ler F_1 WT	23,809	927	8,144	12,262	449	5.23	2.82	
Col/Ler F_1 WT	25,445	909	8,008	13,107	409	4.86	2.63	
Col/Ler F_1 WT	23,774	1,518	6,345	13,186	390	5.79	3.13	
Col/Ler F_1 WT	23,064	863	8,141	11,788	422	4.93	2.66	
Col/Ler F_1 WT	22,946	951	8,074	12,150	462	5.41	2.93	
Col/Ler F_1 WT	21,778	1,978	4,844	11,616	332	6.41	3.47	
Total	**220,967**	**10,288**	**68,191**	**115,780**	**3,875**	**5.44**	**2.94**	**-**
Col/Ler F_1 *msh2*	25,505	951	7,419	12,825	504	6.36	3.44	
Col/Ler F_1 *msh2*	21,105	1,271	5,877	11,013	440	6.97	3.77	
Col/Ler F_1 *msh2*	23,970	932	6,957	11,576	519	6.94	3.75	
Col/Ler F_1 *msh2*	29,605	1,138	8,073	14,903	601	6.93	3.75	
Col/Ler F_1 *msh2*	28,047	1,287	8,186	14,304	612	6.96	3.76	
Col/Ler F_1 *msh2*	24,880	1,220	7,980	13,162	594	6.93	3.74	
Col/Ler F_1 *msh2*	20,604	931	5,701	10,550	475	7.69	4.16	
Col/Ler F_1 *msh2*	28,238	1,140	7,932	13,949	575	6.76	3.65	
Col/Ler F_1 *msh2*	24,971	995	8,056	12,715	529	6.16	3.33	
Total	**226,925**	**9,865**	**66,181**	**114,997**	**4,849**	**6.85**	**3.71**	**0.0005**
Col/CLC F_1 WT	22,353	1,225	7,975	12,429	724	8.32	4.50	
Col/CLC F_1 WT	25,381	1,426	8,156	15,041	758	8.50	4.60	
Col/CLC F_1 WT	16,356	1,153	8,480	5,711	1,012	10.66	5.76	
Col/CLC F_1 WT	5,183	373	1,489	3,164	157	9.54	5.16	
Col/CLC F_1 WT	22,224	1,233	8,014	12,318	659	7.60	4.11	
Col/CLC F_1 WT	16,659	1,434	8,226	6,032	967	10.52	5.69	
Col/CLC F_1 WT	15,069	1,181	8,022	5,174	692	7.94	4.29	
Col/CLC F_1 WT	14,673	902	3,251	10,196	324	9.06	4.90	
Total	**137,898**	**8,927**	**53,613**	**70,065**	**5,293**	**9.02**	**4.87**	**-**
Col/CLC F_1 *msh2*	19,658	1,740	9,000	7,020	1,898	17.42	9.41	
Col/CLC F_1 *msh2*	18,586	1,639	8,117	7,079	1,751	17.74	9.59	
Col/CLC F_1 *msh2*	18,643	1,618	8,073	7,305	1,647	16.94	9.16	
Col/CLC F_1 *msh2*	12,776	935	5,029	5,470	1,342	21.06	11.39	
Col/CLC F_1 *msh2*	21,749	1,775	8,153	9,993	1,828	18.31	9.90	
Col/CLC F_1 *msh2*	19,072	1,567	8,046	7,617	1,842	18.63	10.07	
Total	**110,484**	**9,274**	**46,418**	**44,484**	**10,308**	**18.35**	**9.92**	**0.0007**
Col/Ct F_1 WT	5,180	428	1,636	2,972	144	8.09	4.37	
Col/Ct F_1 WT	23,587	2,918	6,074	14,190	405	6.25	3.38	

Col/Ct F_1 WT	10,741	1,196	3,361	5,991	193	5.43	2.94	
Col/Ct F_1 WT	5,349	568	1,502	3,155	124	7.63	4.12	
Col/Ct F_1 WT	5,204	514	1,520	3,037	133	8.05	4.35	
Col/Ct F_1 WT	36,335	3,648	10,010	21,861	816	7.54	4.07	
Col/Ct F_1 WT	12,116	1,269	3,510	7,085	252	6.70	3.62	
Col/Ct F_1 WT	12,317	1,193	3,510	7,351	263	6.97	3.77	
Col/Ct F_1 WT	11,676	1,632	2,972	6,858	214	6.72	3.63	
Total	**122,505**	**13,366**	**34,095**	**72,500**	**2,544**	**7.04**	**3.81**	**-**
Col/Ct F_1 *msh2*	18,459	2,424	4,706	10,903	426	8.30	4.49	
Col/Ct F_1 *msh2*	5,441	467	1,505	3,345	124	7.61	4.11	
Col/Ct F_1 *msh2*	11,055	1,014	3,018	6,749	274	8.32	4.50	
Col/Ct F_1 *msh2*	8,336	1,021	1,810	5,361	144	7.37	3.98	
Col/Ct F_1 *msh2*	8,215	1,039	2,120	4,878	178	7.75	4.19	
Col/Ct F_1 *msh2*	5,242	551	1,661	2,923	107	6.05	3.27	
Col/Ct F_1 *msh2*	5,406	607	1,424	3,267	108	7.05	3.81	
Col/Ct F_1 *msh2*	10,578	1,159	3,000	6,190	229	7.09	3.83	
Total	**72,732**	**8,282**	**19,244**	**43,616**	**1,590**	**7.44**	**4.02**	**0.321**

Appendix 7.5 – *I2f* flow cytometry fluorescent pollen count data for the wild type and *msh2-1* mutant in hybrid backgrounds. Genetic distance in centimorgans (cM) is calculated as cM = $100\times(N_Y/(N_Y+N_{R+Y}))$, where N_Y is the number of yellow alone pollen, and N_{Y+R} is the number of red and yellow pollen. Mann-Whitney *U* tests were performed to test for significant differences between genotypes.

Individual	Total	Red alone	Red + Yellow	Non-colour	Yellow alone	cM	cM/Mb	*P*-value
Col/Ler F_1 WT	26,874	2,629	8,034	15,655	556	6.47	9.66	
Col/Ler F_1 WT	23,210	1,559	7,766	13,269	616	7.35	10.97	
Col/Ler F_1 WT	24,299	1,655	8,094	13,950	600	6.90	10.30	
Col/Ler F_1 WT	24,601	1,417	8,018	14,711	455	5.37	8.01	
Col/Ler F_1 WT	10,420	1,106	2,996	6,108	210	6.55	9.78	
Col/Ler F_1 WT	23,716	1,709	8,066	13,370	571	6.61	9.87	
Col/Ler F_1 WT	25,297	2,062	8,055	14,541	639	7.35	10.97	
Total	**158,417**	**12,137**	**51,029**	**91,604**	**3,647**	**6.66**	**9.94**	**-**
Col/Ler F_1 *msh2*	11,899	1,029	3,365	7,169	336	9.08	13.55	
Col/Ler F_1 *msh2*	26,006	1,682	7,679	15,934	711	8.47	12.65	
Col/Ler F_1 *msh2*	17,612	1,102	4,976	11,003	531	9.64	14.39	
Col/Ler F_1 *msh2*	23,969	2,354	6,306	14,733	576	8.37	12.49	
Col/Ler F_1 *msh2*	21,379	1,378	7,069	12,366	566	7.41	11.06	
Total	**100,865**	**7,545**	**29,395**	**61,205**	**2,720**	**8.60**	**12.83**	**0.006**
Col/CLC F_1 WT	19,279	1,372	8,027	8,998	882	9.90	14.78	
Col/CLC F_1 WT	12,282	1,592	8,471	1,383	836	8.98	13.41	
Col/CLC F_1 WT	16,415	1,397	8,002	6,082	934	10.45	15.60	
Col/CLC F_1 WT	17,366	489	7,926	8,163	788	9.04	13.50	
Col/CLC F_1 WT	21,144	1,106	8,193	10,954	891	9.81	14.64	
Col/CLC F_1 WT	17,099	1,583	7,524	7,094	898	10.66	15.91	
Total	**103,585**	**7,539**	**48,143**	**42,674**	**5,229**	**9.81**	**14.64**	**-**
Col/CLC F_1 *msh2*	18,205	1,655	6,986	8,017	1,547	18.13	27.06	
Col/CLC F_1 *msh2*	18,169	1,573	7,928	7,265	1,403	15.04	22.44	
Col/CLC F_1 *msh2*	18,362	1,521	8,002	7,198	1,641	17.02	25.40	
Col/CLC F_1 *msh2*	13,593	867	7,267	4,098	1,361	15.77	23.54	
Col/CLC F_1 *msh2*	21,599	1,578	8,274	9,967	1,780	17.70	26.42	
Total	**89,928**	**7,194**	**38,457**	**36,545**	**7,732**	**16.73**	**24.97**	**0.004**

Appendix 7.6 – *5.10* wild type and *msh2-1* mutant F_2 fluorescent seed count data. Genetic distance in centimorgans (cM) is calculated as $cM = 100\times(1-(1-2(N_G+N_R)/N_T)^{1/2})$, where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds of all classes analysed. Mann-Whitney *U* tests were performed to test for significant differences between genotypes.

Individual	Green alone	Red alone	Red and Green	Non-colour	Total seed	cM	cM/ Mb	Green: Non-green	Red: Non-red	Green alone: Red alone
Col-0 WT	195	181	1,392	342	2,110	19.78	2.95	3.03	2.93	1.08
Col-0 WT	106	81	663	164	1,014	20.55	3.07	3.14	2.76	1.31
Col-0 WT	155	119	1,052	281	1,607	18.82	2.81	3.02	2.69	1.30
Col-0 WT	166	186	1,209	248	1,809	21.84	3.26	3.17	3.37	0.89
Col-0 WT	165	147	1,178	293	1,783	19.38	2.89	3.05	2.89	1.12
Col-0 WT	169	177	1,359	321	2,026	18.86	2.81	3.07	3.13	0.95
Total	**956**	**891**	**6,853**	**1,649**	**10,349**	**19.87**	**2.97**	**-**		
Col-0 *msh2*	195	185	1,319	319	2,018	21.04	3.14	3.00	2.93	1.05
Col-0 *msh2*	95	99	664	201	1,059	20.40	3.04	2.53	2.58	0.96
Col-0 *msh2*	123	128	982	218	1,451	19.13	2.85	3.19	3.26	0.96
Col-0 *msh2*	156	144	1,207	327	1,834	17.97	2.68	2.89	2.80	1.08
Col-0 *msh2*	156	135	1,070	246	1,607	20.14	3.01	3.22	3.00	1.16
Col-0 *msh2*	176	186	1,373	390	2,125	18.80	2.81	2.69	2.75	0.95
Col-0 *msh2*	180	176	1,425	349	2,130	18.41	2.75	3.06	3.03	1.02
Col-0 *msh2*	175	175	1,362	347	2,059	18.76	2.80	2.94	2.94	1.00
Col-0 *msh2*	168	171	1,191	324	1,854	20.36	3.04	2.75	2.77	0.98
Col-0 *msh2*	186	192	1,401	339	2,118	19.81	2.96	2.99	3.03	0.97
Total	**1,610**	**1,591**	**11,994**	**3,060**	**18,255**	**19.48**	**2.91**	***P* = 0.492**		
Col/Ler F_1 WT	187	196	1,413	361	2,157	19.70	2.94	2.87	2.94	0.95
Col/Ler F_1 WT	155	155	1,342	297	1,949	17.42	2.60	3.31	3.31	1.00
Col/Ler F_1 WT	190	180	1,322	291	1,983	20.83	3.11	3.21	3.12	1.06
Col/Ler F_1 WT	180	183	1,364	336	2,063	19.50	2.91	2.97	3.00	0.98
Col/Ler F_1 WT	148	158	1,146	291	1,743	19.45	2.90	2.88	2.97	0.94
Col/Ler F_1 WT	211	222	1,482	327	2,242	21.66	3.23	3.08	3.17	0.95
Col/Ler F_1 WT	173	202	1,424	368	2,167	19.14	2.86	2.80	3.01	0.86
Col/Ler F_1 WT	190	194	1,417	339	2,140	19.93	2.97	3.02	3.05	0.98
Col/Ler F_1 WT	179	185	1,413	307	2,084	19.34	2.89	3.24	3.29	0.97
Col/Ler F_1 WT	176	186	1,377	385	2,124	18.81	2.81	2.72	2.79	0.95
Total	**1,789**	**1,861**	**13,700**	**3,302**	**20,652**	**19.58**	**2.92**	**-**		
Col/Ler F_1 *msh2*	147	158	1,207	292	1,804	18.65	2.78	3.01	3.11	0.93
Col/Ler F_1 *msh2*	159	175	1,441	344	2,119	17.25	2.57	3.08	3.21	0.91
Col/Ler F_1 *msh2*	162	168	1,297	318	1,945	18.72	2.79	3.00	3.05	0.96
Col/Ler F_1 *msh2*	156	154	1,345	343	1,998	16.95	2.53	3.02	3.00	1.01
Col/Ler F_1 *msh2*	160	171	1,383	372	2,086	17.38	2.59	2.84	2.92	0.94
Col/Ler F_1 *msh2*	172	194	1,315	322	2,003	20.34	3.04	2.88	3.05	0.89
Col/Ler F_1 *msh2*	146	109	1,215	271	1,741	15.91	2.38	3.58	3.18	1.34
Total	**1,102**	**1,129**	**9,203**	**2,262**	**13,696**	**17.89**	**2.67**	***P* = 0.0136**		
Col/CLC F_1 WT	231	219	1,341	253	2,044	25.19	3.76	3.33	3.22	1.05
Col/CLC F_1 WT	97	95	532	122	846	26.10	3.90	2.90	2.86	1.02

Col/CLC F_1 WT	199	195	1,251	251	1,896	23.55	3.52	3.25	3.21	1.02
Col/CLC F_1 WT	204	220	1,261	278	1,963	24.63	3.68	2.94	3.07	0.93
Col/CLC F_1 WT	191	191	1,149	237	1,768	24.64	3.68	3.13	3.13	1.00
Col/CLC F_1 WT	233	225	1,273	282	2,013	26.18	3.91	2.97	2.91	1.04
Col/CLC F_1 WT	252	251	1,361	297	2,161	26.89	4.01	2.94	2.94	1.00
Col/CLC F_1 WT	217	236	1,379	305	2,137	24.10	3.60	2.95	3.09	0.92
Col/CLC F_1 WT	218	229	1,364	292	2,103	24.18	3.61	3.04	3.12	0.95
Col/CLC F_1 WT	209	236	1,350	299	2,094	24.17	3.61	2.91	3.12	0.89
Col/CLC F_1 WT	236	228	1,297	255	2,016	26.54	3.96	3.17	3.11	1.04
Col/CLC F_1 WT	231	245	1,245	275	1,996	27.68	4.13	2.84	2.94	0.94
Col/CLC F_1 WT	248	241	1,385	291	2,165	25.95	3.87	3.07	3.02	1.03
Col/CLC F_1 WT	218	196	1,164	266	1,844	25.77	3.85	2.99	2.81	1.11
Col/CLC F_1 WT	230	209	1,206	258	1,903	26.61	3.97	3.07	2.90	1.10
Col/CLC F_1 WT	170	154	1,016	209	1,549	23.73	3.54	3.27	3.09	1.10
Col/CLC F_1 WT	211	284	1,373	282	2,150	26.55	3.96	2.80	3.36	0.74
Total	**3,595**	**3,654**	**20,947**	**4,452**	**32,648**	**25.44**	**3.80**	**-**		
Col/CLC F_1 *msh2*	216	212	1,283	268	1,979	24.67	3.68	3.12	3.09	1.02
Col/CLC F_1 *msh2*	236	236	1,350	277	2,099	25.82	3.85	3.09	3.09	1.00
Col/CLC F_1 *msh2*	147	152	992	220	1,511	22.27	3.32	3.06	3.12	0.97
Col/CLC F_1 *msh2*	233	242	1,387	288	2,150	25.29	3.77	3.06	3.13	0.96
Col/CLC F_1 *msh2*	244	240	1,406	287	2,177	25.48	3.80	3.13	3.10	1.02
Col/CLC F_1 *msh2*	215	222	1,298	298	2,033	24.50	3.66	2.91	2.96	0.97
Col/CLC F_1 *msh2*	245	231	1,451	338	2,265	23.86	3.56	2.98	2.89	1.06
Col/CLC F_1 *msh2*	202	184	1,182	297	1,865	23.45	3.50	2.88	2.74	1.10
Col/CLC F_1 *msh2*	192	211	1,282	278	1,963	23.23	3.47	3.01	3.18	0.91
Total	**1,930**	**1,930**	**11,631**	**2,551**	**18,042**	**24.28**	**3.62**	***P* = 0.0446**		
Col/Ct F_1 WT	203	184	1,271	316	1,974	22.03	3.29	2.95	2.80	1.10
Col/Ct F_1 WT	184	200	1,340	309	2,033	21.12	3.15	2.99	3.12	0.92
Col/Ct F_1 WT	180	172	1,107	309	1,768	22.42	3.35	2.68	2.62	1.05
Col/Ct F_1 WT	198	191	1,128	274	1,791	24.79	3.70	2.85	2.79	1.04
Col/Ct F_1 WT	191	189	1,374	305	2,059	20.57	3.07	3.17	3.15	1.01
Col/Ct F_1 WT	177	192	1,256	318	1,943	21.25	3.17	2.81	2.93	0.92
Col/Ct F_1 WT	147	152	1,081	238	1,618	20.60	3.07	3.15	3.20	0.97
Col/Ct F_1 WT	207	184	1,315	311	2,017	21.75	3.25	3.07	2.89	1.13
Total	**1,487**	**1,464**	**9,872**	**2,380**	**15,203**	**21.82**	**3.26**	**-**		
Col/Ct F_1 *msh2*	139	146	1,080	276	1,641	19.21	2.87	2.89	2.95	0.95
Col/Ct F_1 *msh2*	178	195	1,275	325	1,973	21.14	3.16	2.79	2.92	0.91
Col/Ct F_1 *msh2*	181	174	1,214	247	1,816	21.96	3.28	3.31	3.24	1.04
Col/Ct F_1 *msh2*	182	183	1,413	350	2,128	18.95	2.83	2.99	3.00	0.99
Col/Ct F_1 *msh2*	153	147	1,182	278	1,760	18.82	2.81	3.14	3.08	1.04
Col/Ct F_1 *msh2*	218	200	1,349	347	2,114	22.25	3.32	2.86	2.74	1.09
Col/Ct F_1 *msh2*	188	169	1,342	321	2,020	19.59	2.92	3.12	2.97	1.11
Col/Ct F_1 *msh2*	145	156	1,170	265	1,736	19.18	2.86	3.12	3.23	0.93
Col/Ct F_1 *msh2*	162	150	1,218	330	1,860	18.48	2.76	2.88	2.78	1.08
Total	**1,546**	**1,520**	**11,243**	**2,739**	**17,048**	**19.95**	**2.98**	***P* = 0.0360**		

Appendix 7.7 – *5.11* wild type and *msh2-1* mutant F_2 fluorescent seed count data. Genetic distance in centimorgans (cM) is calculated as $cM = 100\times(1-(1-2(N_G+N_R)/N_T)^{1/2})$, where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds of all classes analysed. Mann-Whitney *U* tests were performed to test for significant differences between genotypes.

Individual	Green alone	Red alone	Red and Green	Non-colour	Total seed	cM	cM/ Mb	Green: Non-green	Red: Non-red	Green alone: Red alone
Col-0 WT	163	173	1,205	317	1,858	20.11	2.87	2.79	2.87	0.94
Col-0 WT	164	195	1,360	308	2,027	19.64	2.81	3.03	3.29	0.84
Col-0 WT	173	197	1,409	371	2,150	19.02	2.72	2.79	2.95	0.88
Col-0 WT	222	185	1,469	325	2,201	20.62	2.95	3.32	3.02	1.20
Col-0 WT	188	184	1,354	331	2,057	20.11	2.87	2.99	2.96	1.02
Col-0 WT	218	202	1,439	354	2,213	21.23	3.03	2.98	2.87	1.08
Col-0 WT	216	168	1,479	388	2,251	18.83	2.69	3.05	2.73	1.29
Col-0 WT	114	109	873	221	1,317	18.68	2.67	2.99	2.93	1.05
Col-0 WT	194	191	1,417	307	2,109	20.32	2.90	3.23	3.21	1.02
Col-0 WT	135	149	1,092	341	1,717	18.20	2.60	2.50	2.61	0.91
Col-0 WT	151	147	1,186	277	1,761	18.66	2.67	3.15	3.11	1.03
Col-0 WT	211	192	1,454	352	2,209	20.31	2.90	3.06	2.92	1.10
Col-0 WT	200	186	1,517	367	2,270	18.77	2.68	3.10	3.00	1.08
Col-0 WT	182	200	1,461	316	2,159	19.62	2.80	3.18	3.34	0.91
Total	**2,531**	**2,478**	**18,715**	**4,575**	**28,299**	**19.58**	**2.80**	**-**		
Col-0 *msh2*	186	186	1,408	363	2,143	19.20	2.74	2.90	2.90	1.00
Col-0 *msh2*	162	160	1,331	327	1,980	17.86	2.55	3.07	3.05	1.01
Col-0 *msh2*	158	148	1,138	315	1,759	19.25	2.75	2.80	2.72	1.07
Col-0 *msh2*	208	180	1,313	313	2,014	21.60	3.09	3.09	2.87	1.16
Col-0 *msh2*	144	162	1,257	326	1,889	17.78	2.54	2.87	3.02	0.89
Col-0 *msh2*	186	211	1,464	384	2,245	19.61	2.80	2.77	2.94	0.88
Col-0 *msh2*	168	184	1,271	332	1,955	20.01	2.86	2.79	2.91	0.91
Col-0 *msh2*	171	139	1,201	241	1,752	19.62	2.80	3.61	3.25	1.23
Col-0 *msh2*	151	179	1,245	364	1,939	18.78	2.68	2.57	2.77	0.84
Col-0 *msh2*	165	147	1,178	293	1,783	19.38	2.77	3.05	2.89	1.12
Col-0 *msh2*	165	147	1,178	293	1,783	19.38	2.77	3.05	2.89	1.12
Col-0 *msh2*	165	147	1,178	293	1,783	19.38	2.77	3.05	2.89	1.12
Col-0 *msh2*	169	182	1,342	300	1,993	19.52	2.79	3.13	3.25	0.93
Col-0 *msh2*	178	197	1,390	338	2,103	19.79	2.83	2.93	3.08	0.90
Total	**2,376**	**2,369**	**17,894**	**4,482**	**27,121**	**19.37**	**2.77**	***P* = 0.490**		
Col/Ler F_1 WT	146	152	962	261	1,521	22.02	3.15	2.68	2.74	0.96
Col/Ler F_1 WT	121	115	903	200	1,339	19.53	2.79	3.25	3.17	1.05
Col/Ler F_1 WT	195	188	1,282	309	1,974	21.77	3.11	2.97	2.92	1.04
Col/Ler F_1 WT	208	211	1,434	334	2,187	21.46	3.07	3.01	3.04	0.99
Col/Ler F_1 WT	199	209	1,586	397	2,391	18.84	2.69	2.95	3.01	0.95
Col/Ler F_1 WT	193	215	1,471	399	2,278	19.89	2.84	2.71	2.85	0.90
Col/Ler F_1 WT	228	223	1,554	360	2,365	21.35	3.05	3.06	3.02	1.02
Col/Ler F_1 WT	202	204	1,381	350	2,137	21.26	3.04	2.86	2.87	0.99
Col/Ler F_1 WT	196	205	1,358	348	2,107	21.30	3.04	2.81	2.87	0.96

Col/Ler F_1 WT	177	199	1,368	334	2,078	20.12	2.87	2.90	3.07	0.89
Col/Ler F_1 WT	205	217	1,399	357	2,178	21.74	3.11	2.79	2.88	0.94
Col/Ler F_1 WT	181	190	1,283	307	1,961	21.16	3.02	2.95	3.02	0.95
Col/Ler F_1 WT	192	225	1,395	307	2,119	22.13	3.16	2.98	3.25	0.85
Total	**2,443**	**2,553**	**17,376**	**4,263**	**26,635**	**20.97**	**3.00**	**-**		
Col/Ler F_1 *msh2*	186	197	1,542	356	2,281	18.50	2.64	3.12	3.21	0.94
Col/Ler F_1 *msh2*	138	145	1,111	261	1,655	18.88	2.70	3.08	3.15	0.95
Col/Ler F_1 *msh2*	120	121	900	225	1,366	19.55	2.79	2.95	2.96	0.99
Col/Ler F_1 *msh2*	197	208	1,427	384	2,216	20.35	2.91	2.74	2.81	0.95
Col/Ler F_1 *msh2*	178	175	1,427	383	2,163	17.93	2.56	2.88	2.86	1.02
Col/Ler F_1 *msh2*	159	155	988	234	1,536	23.11	3.30	2.95	2.91	1.03
Col/Ler F_1 *msh2*	199	186	1,423	377	2,185	19.53	2.79	2.88	2.79	1.07
Col/Ler F_1 *msh2*	186	201	1,469	374	2,230	19.20	2.74	2.88	2.98	0.93
Col/Ler F_1 *msh2*	177	192	1,521	360	2,250	18.02	2.57	3.08	3.19	0.92
Total	**1,540**	**1,580**	**11,808**	**2,954**	**17,882**	**19.45**	**2.78**	***P* = 0.0177**		
Col/CLC F_1 WT	240	225	1,234	290	1,989	27.03	3.86	2.86	2.75	1.07
Col/CLC F_1 WT	259	261	1,422	303	2,245	26.74	3.82	2.98	2.99	0.99
Col/CLC F_1 WT	247	243	1,230	261	1,981	28.92	4.13	2.93	2.90	1.02
Total	**746**	**729**	**3,886**	**854**	**6,215**	**27.56**	**3.94**	**-**		
Col/CLC F_1 *msh2*	194	210	1,343	259	2,006	22.72	3.25	3.28	3.43	0.92
Col/CLC F_1 *msh2*	147	177	991	207	1,522	24.22	3.46	2.96	3.30	0.83
Col/CLC F_1 *msh2*	190	190	1,203	269	1,852	23.21	3.32	3.03	3.03	1.00
Col/CLC F_1 *msh2*	206	214	1,258	267	1,945	24.63	3.52	3.04	3.11	0.96
Col/CLC F_1 *msh2*	200	195	1,213	230	1,838	24.49	3.50	3.32	3.27	1.03
Total	**937**	**986**	**6,008**	**1,232**	**9,163**	**23.85**	**3.41**	***P* = 0.0357**		
Col/Ct F_1 WT	197	221	1,402	311	2,131	22.05	3.15	3.01	3.19	0.89
Col/Ct F_1 WT	257	244	1,334	326	2,161	26.77	3.82	2.79	2.71	1.05
Col/Ct F_1 WT	205	198	1,324	302	2,029	22.36	3.19	3.06	3.00	1.04
Col/Ct F_1 WT	223	189	1,349	303	2,064	22.49	3.21	3.20	2.92	1.18
Col/Ct F_1 WT	265	268	1,431	275	2,239	27.62	3.95	3.12	3.15	0.99
Col/Ct F_1 WT	205	198	1,271	261	1,935	23.62	3.37	3.22	3.15	1.04
Col/Ct F_1 WT	167	165	1,235	277	1,844	20.01	2.86	3.17	3.15	1.01
Col/Ct F_1 WT	187	216	1,262	248	1,913	23.93	3.42	3.12	3.40	0.87
Col/Ct F_1 WT	188	217	1,071	205	1,681	28.02	4.00	2.98	3.28	0.87
Total	**1,894**	**1,916**	**11,679**	**2,508**	**17,997**	**24.09**	**3.44**	**-**		
Col/Ct F_1 *msh2*	134	145	1,164	284	1,727	17.73	2.53	3.03	3.13	0.92
Col/Ct F_1 *msh2*	238	260	1,471	362	2,331	24.32	3.47	2.75	2.89	0.92
Col/Ct F_1 *msh2*	153	145	1,256	283	1,837	17.81	2.54	3.29	3.21	1.06
Col/Ct F_1 *msh2*	152	147	1,115	295	1,709	19.37	2.77	2.87	2.82	1.03
Col/Ct F_1 *msh2*	154	159	1,243	302	1,858	18.57	2.65	3.03	3.07	0.97
Col/Ct F_1 *msh2*	260	241	1,386	315	2,202	26.18	3.74	2.96	2.83	1.08
Col/Ct F_1 *msh2*	266	301	1,514	321	2,402	27.34	3.91	2.86	3.09	0.88
Col/Ct F_1 *msh2*	234	262	1,337	282	2,115	27.13	3.88	2.89	3.10	0.89
Total	**1,591**	**1,660**	**10,486**	**2,444**	**16,181**	**22.31**	**3.19**	***P* = 0.370**		

Appendix 7.8 – Alexander staining pollen viability counts in wild type and *msh2-1* mutants, in inbred and hybrid backgrounds. Pollen viability assays were performed on an average of nine plants per genotype. Per sample, three mature flowers were selected from the primary floral axis, and pollen was pooled during collection. Viable pollen grains stain purple, whereas inviable pollen is misshapen and stains blue-green. To test for statistically significant differences between wild type and *msh2-1*, 2×2 contingency tables were constructed using aggregate viable and inviable pollen and X^2 tests performed.

Individual	Viable	Inviable	Total	% viable	% inviable	*P*-value
Col-0 WT	1,498	24	1,522	98.4	1.6	
Col-0 WT	1,484	26	1,510	98.3	1.7	
Col-0 WT	1,712	16	1,728	99.1	0.9	
Col-0 WT	1,591	14	1,605	99.1	0.9	
Col-0 WT	1,837	20	1,857	98.9	1.1	
Col-0 WT	1,793	15	1,808	99.2	0.8	
Col-0 WT	1,972	27	1,999	98.6	1.4	
Total	**11,887**	**142**	**12,029**	**98.8**	**1.2**	**-**
Col-0 *msh2*	1,464	36	1,500	97.6	2.4	
Col-0 *msh2*	1,398	40	1,438	97.2	2.8	
Col-0 *msh2*	1,361	26	1,387	98.1	1.9	
Col-0 *msh2*	1,683	47	1,730	97.3	2.7	
Col-0 *msh2*	1,831	62	1,893	96.7	3.3	
Col-0 *msh2*	1,892	39	1,931	98.0	2.0	
Col-0 *msh2*	1,995	23	2,018	98.9	1.1	
Total	**11,624**	**273**	**11,897**	**97.7**	**2.3**	**5.7×10^{-11}**
Ler-0 WT	1,420	55	1,475	96.3	3.7	
Ler-0 WT	2,000	4	2,004	99.8	0.2	
Ler-0 WT	2,089	5	2,094	99.8	0.2	
Ler-0 WT	2,147	11	2,158	99.5	0.5	
Ler-0 WT	2,016	22	2,038	98.9	1.1	
Ler-0 WT	2,016	6	2,022	99.7	0.3	
Ler-0 WT	2,015	11	2,026	99.5	0.5	
Ler-0 WT	2,135	8	2,143	99.6	0.4	
Ler-0 WT	2,038	7	2,045	99.7	0.3	
Ler-0 WT	2,012	14	2,026	99.3	0.7	
Ler-0 WT	2,009	7	2,016	99.7	0.3	
Total	**21,897**	**150**	**22,047**	**99.2**	**0.8**	**-**
Ler-0 *msh2*	2,309	65	2,374	97.3	2.7	
Ler-0 *msh2*	2,032	50	2,082	97.6	2.4	
Ler-0 *msh2*	2,043	85	2,128	96.0	4.0	
Ler-0 *msh2*	921	37	958	96.1	3.9	
Ler-0 *msh2*	2,032	61	2,093	97.1	2.9	
Ler-0 *msh2*	2,004	128	2,132	94.0	6.0	
Ler-0 *msh2*	2,031	41	2,072	98.0	2.0	
Ler-0 *msh2*	2,061	40	2,101	98.1	1.9	
Ler-0 *msh2*	1,981	179	2,160	91.7	8.3	

Total	**17,414**	**686**	**18,100**	**96.2**	**3.8**	**< 2.2×10^{-16}**
CLC WT	1,357	6	1,363	99.6	0.4	
CLC WT	1,343	11	1,354	99.2	0.8	
CLC WT	1,612	3	1,615	99.8	0.2	
CLC WT	1,629	3	1,632	99.8	0.2	
CLC WT	1,642	7	1,649	99.6	0.4	
CLC WT	1,346	9	1,355	99.3	0.7	
CLC WT	1,706	12	1,718	99.3	0.7	
CLC WT	1,199	13	1,212	98.9	1.1	
CLC WT	1,636	19	1,655	98.9	1.1	
CLC WT	1,206	9	1,215	99.3	0.7	
CLC WT	1,721	14	1,735	99.2	0.8	
Total	**16,397**	**106**	**16,503**	**99.3**	**0.7**	**-**
CLC *msh2*	1,866	29	1,895	98.5	1.5	
CLC *msh2*	1,841	39	1,880	97.9	2.1	
CLC *msh2*	1,866	9	1,875	99.5	0.5	
CLC *msh2*	1,917	13	1,930	99.3	0.7	
CLC *msh2*	1,835	85	1,920	95.6	4.4	
CLC *msh2*	1,406	15	1,421	98.9	1.1	
CLC *msh2*	1,938	14	1,952	99.3	0.7	
CLC *msh2*	1,739	221	1,960	88.7	11.3	
CLC *msh2*	1,906	114	2,020	94.4	5.6	
Total	**16,314**	**539**	**16,853**	**96.9**	**3.1**	**< 2.2×10^{-16}**
Col/CLC F_1 WT	2,054	40	2,094	98.1	1.9	
Col/CLC F_1 WT	1,857	18	1,875	99.0	1.0	
Col/CLC F_1 WT	1,950	47	1,997	97.6	2.4	
Col/CLC F_1 WT	2,025	35	2,060	98.3	1.7	
Col/CLC F_1 WT	2,000	51	2,051	97.5	2.5	
Col/CLC F_1 WT	2,046	56	2,102	97.3	2.7	
Col/CLC F_1 WT	1,995	23	2,018	98.9	1.1	
Col/CLC F_1 WT	1,462	23	1,485	98.5	1.5	
Total	**15,389**	**293**	**15,682**	**98.2**	**1.8**	**-**
Col/CLC F_1 *msh2*	2,037	44	2,081	97.9	2.1	
Col/CLC F_1 *msh2*	1,836	48	1,884	97.5	2.5	
Col/CLC F_1 *msh2*	1,856	70	1,926	96.4	3.6	
Col/CLC F_1 *msh2*	1,917	34	1,951	98.3	1.7	
Col/CLC F_1 *msh2*	2,076	56	2,132	97.4	2.6	
Col/CLC F_1 *msh2*	2,088	50	2,138	97.7	2.3	
Col/CLC F_1 *msh2*	1,997	49	2,046	97.6	2.4	
Col/CLC F_1 *msh2*	2,212	73	2,285	96.8	3.2	
Col/CLC F_1 *msh2*	2,032	117	2,149	94.6	5.4	
Col/CLC F_1 *msh2*	2,031	57	2,088	97.3	2.7	
Col/CLC F_1 *msh2*	2,178	42	2,220	98.1	1.9	
Col/CLC F_1 *msh2*	1,968	28	1,996	98.6	1.4	
Total	**24,228**	**668**	**24,896**	**97.3**	**2.7**	**5.18×10^{-8}**

Appendix 7.9 – Proportion of rod and ring bivalents at metaphase I in DAPI spreads from hybrid *msh2-1* and wild type. The five bivalents of metaphase I DAPI spreads were scored as either rod (1 chiasma) or ring (2+ chiasma), depending on chromosome morphology (López et al., 2012). Statistical analysis of these data is presented in appendix 7.10.

Individual	No. 'rod'	No. 'ring'	Prop. 'rod'	Prop. 'ring'
Col/Ler F_1 WT	3	2	0.6	0.4
Col/Ler F_1 WT	2	3	0.4	0.6
Col/Ler F_1 WT	2	3	0.4	0.6
Col/Ler F_1 WT	2	3	0.4	0.6
Col/Ler F_1 WT	4	1	0.8	0.2
Col/Ler F_1 WT	2	3	0.4	0.6
Col/Ler F_1 WT	4	1	0.8	0.2
Col/Ler F_1 WT	2	3	0.4	0.6
Col/Ler F_1 WT	2	3	0.4	0.6
Col/Ler F_1 WT	3	2	0.6	0.4
Col/Ler F_1 WT	2	3	0.4	0.6
Col/Ler F_1 WT	2	3	0.4	0.6
Col/Ler F_1 WT	4	1	0.8	0.2
Col/Ler F_1 WT	2	3	0.4	0.6
Col/Ler F_1 WT	3	2	0.6	0.4
Col/Ler F_1 WT	3	2	0.6	0.4
Col/Ler F_1 WT	3	2	0.6	0.4
Col/Ler F_1 WT	2	3	0.4	0.6
Col/Ler F_1 WT	5	0	1	0
Col/Ler F_1 WT	4	1	0.8	0.2
Total	**56**	**44**	**0.56**	**0.44**
Col/Ler F_1 *msh2*	3	2	0.6	0.4
Col/Ler F_1 *msh2*	3	2	0.6	0.4
Col/Ler F_1 *msh2*	2	3	0.4	0.6
Col/Ler F_1 *msh2*	4	1	0.8	0.2
Col/Ler F_1 *msh2*	4	1	0.8	0.2
Col/Ler F_1 *msh2*	1	4	0.2	0.8
Col/Ler F_1 *msh2*	3	2	0.6	0.4
Col/Ler F_1 *msh2*	2	3	0.4	0.6
Total	**22**	**18**	**0.55**	**0.45**
Col/CLC F_1 WT	1	4	0.2	0.8
Col/CLC F_1 WT	1	4	0.2	0.8
Col/CLC F_1 WT	0	5	0	1
Col/CLC F_1 WT	2	3	0.4	0.6
Col/CLC F_1 WT	0	5	0	1
Col/CLC F_1 WT	0	5	0	1
Col/CLC F_1 WT	2	3	0.4	0.6
Total	**6**	**29**	**0.17**	**0.83**
Col/CLC F_1 *msh2*	1	4	0.2	0.8
Col/CLC F_1 *msh2*	2	3	0.4	0.6

Col/CLC F_1 *msh2*	2	3	0.4	0.6
Col/CLC F_1 *msh2*	1	4	0.2	0.8
Col/CLC F_1 *msh2*	1	4	0.2	0.8
Col/CLC F_1 *msh2*	2	3	0.4	0.6
Total	**9**	**21**	**0.3**	**0.7**

Appendix 7.10 – *P*-value summary from statistical analysis of bivalent 'rod'/'ring' count data. To test for statistically significant differences between genotypes (appendix 7.9), pairwise Wilcoxon rank sum tests were performed using the 'Benjamini & Hochberg' correction for multiple testing.

Genotype	Col/Ler F_1 *msh2*	Col/CLC F_1 WT	Col/CLC F_1 *msh2*
Col/Ler F_1 WT	0.957	0.003	0.011
Col/Ler F_1 *msh2*	-	0.016	0.047
Col/CLC F_1 WT	-	-	0.234

Appendix 7.11 – Read pair number and genome coverage for wild type and *msh2-1* GBS F_2 individuals. The number of GBS read pairs (2 × 151 bp) aligning to the Arabidopsis TAIR10 reference genome is indicated, for both sequencing lanes (96 individuals per lane) of each genotype. Corresponding coverage (Cov) values are indicated. Alignments were performed with Bowtie2. '-' indicates F_2 individuals excluded from the dataset due to a combination of low coverage and/or poor quality allele frequency plots. Col/Ler *msh2-1* data were compared to a previously published set of Col/Ler wild type GBS crossovers (Serra et al., 2018a).

F_2	Col/Ler *msh2-1*				Col/CLC wild type				Col/CLC *msh2-1*			
	Lane 1		Lane2		Lane 1		Lane2		Lane 1		Lane2	
	Read pair #	Cov	Read pair #	Cov	Read pair #	Cov	Read pair #	Cov	Read pair #	Cov	Read pair #	Cov
1	1,590,990	4.03	1,775,754	4.50	3,307,534	8.38	903,415	2.29	1,717,321	4.35	1,564,809	3.97
2	1,714,300	4.35	1,384,604	3.51	1,665,522	4.22	661,285	1.68	2,261,869	5.73	937,471	2.38
3	2,288,781	5.80	1,948,356	4.94	1,975,322	5.01	1,345,572	3.41	1,757,522	4.45	1,818,515	4.61
4	1,625,172	4.12	2,762,147	7.00	2,340,266	5.93	1,155,605	2.93	3,038,373	7.70	1,334,272	3.38
5	2,177,169	5.52	2,735,549	6.93	3,826,430	9.70	1,521,986	3.86	3,157,260	8.00	1,404,982	3.56
6	1,606,500	4.07	2,610,918	6.62	1,077,883	2.73	838,706	2.13	1,895,363	4.80	2,550,778	6.47
7	1,778,405	4.51	2,837,197	7.19	916,338	2.32	1,172,482	2.97	2,784,542	7.06	1,981,400	5.02
8	1,596,016	4.05	1,695,218	4.30	778,019	1.97	1,207,002	3.06	2,929,945	7.43	1,048,850	2.66
9	2,008,480	5.09	1,426,132	3.61	905,611	2.30	572,065	1.45	2,075,097	5.26	1,635,350	4.15
10	1,905,092	4.83	2,420,749	6.14	657,339	1.67	290,344	0.74	1,728,449	4.38	728,681	1.85
11	1,543,443	3.91	2,362,743	5.99	3,596,970	9.12	835,636	2.12	899,933	2.28	1,352,347	3.43
12	2,502,617	6.34	3,270,287	8.29	1,319,630	3.34	968,766	2.46	2,035,805	5.16	2,491,232	6.31
13	1,743,092	4.42	3,422,241	8.67	3,131,213	7.94	1,265,934	3.21	422,688	1.07	2,101,421	5.33
14	1,310,011	3.32	-		2,048,395	5.19	1,039,181	2.63	2,368,911	6.00	1,917,074	4.86
15	2,163,685	5.48	3,136,186	7.95	1,969,791	4.99	1,179,954	2.99	3,421,174	8.67	1,743,037	4.42
16	2,332,595	5.91	1,845,312	4.68	1,857,541	4.71	1,320,141	3.35	1,603,033	4.06	1,433,571	3.63
17	2,519,082	6.39	2,435,717	6.17	1,425,559	3.61	1,117,670	2.83	2,020,899	5.12	1,880,652	4.77
18	1,223,538	3.10	1,289,158	3.27	1,337,728	3.39	611,151	1.55	1,024,240	2.60	986,986	2.50
19	2,049,404	5.19	3,106,677	7.87	1,735,334	4.40	1,207,620	3.06	1,532,347	3.88	1,775,838	4.50
20	1,134,587	2.88	2,475,215	6.27	2,133,963	5.41	1,143,797	2.90	1,781,241	4.51	1,747,916	4.43
21	1,582,878	4.01	2,065,099	5.23	1,639,545	4.16	1,150,411	2.92	799,824	2.03	2,957,119	7.50
22	1,824,574	4.62	3,216,323	8.15	2,039,388	5.17	1,265,727	3.21	1,597,714	4.05	3,222,362	8.17
23	1,593,033	4.04	2,535,664	6.43	2,852,245	7.23	1,086,274	2.75	2,033,187	5.15	2,442,336	6.19
24	1,375,037	3.49	2,984,989	7.57	2,003,962	5.08	1,352,966	3.43	1,334,498	3.38	1,779,859	4.51
25	2,065,452	5.24	2,417,818	6.13	2,083,047	5.28	1,029,383	2.61	1,699,250	4.31	1,548,062	3.92
26	2,175,509	5.51	1,769,361	4.48	2,375,284	6.02	1,449,645	3.67	1,915,628	4.86	1,750,132	4.44
27	1,682,383	4.26	1,535,684	3.89	2,804,307	7.11	1,169,425	2.96	1,385,360	3.51	1,340,170	3.40
28	2,878,525	7.30	2,290,806	5.81	1,336,150	3.39	1,209,405	3.07	1,632,796	4.14	1,714,505	4.35
29	2,504,507	6.35	3,138,381	7.95	1,539,279	3.90	2,259,958	5.73	1,287,441	3.26	1,792,078	4.54
30	950,120	2.41	2,194,506	5.56	2,360,088	5.98	835,476	2.12	1,878,143	4.76	1,769,269	4.48
31	564,198	1.43	1,172,261	2.97	2,280,038	5.78	1,017,073	2.58	2,504,507	6.35	1,313,858	3.33
32	870,780	2.21	898,520	2.28	2,445,316	6.20	1,286,018	3.26	1,932,630	4.90	148,526	0.38
33	1,309,624	3.32	1,223,935	3.10	1,165,070	2.95	824,656	2.09	2,401,394	6.09	1,109,541	2.81
34	2,649,963	6.72	1,680,948	4.26	1,419,631	3.60	-		2,143,191	5.43	1,949,487	4.94
35	1,836,559	4.66	1,041,457	2.64	1,945,609	4.93	-		2,083,568	5.28	1,181,453	2.99
36	1,973,174	5.00	1,463,346	3.71	1,942,030	4.92	1,569,814	3.98	2,556,142	6.48	2,557,121	6.48
37	2,018,862	5.12	1,570,619	3.98	2,393,869	6.07	693,480	1.76	2,188,910	5.55	1,724,416	4.37

38	1,465,608	3.71	2,342,056	5.94	1,719,582	4.36	1,891,777	4.80	1,024,843	2.60	3,121,388	7.91
39	857,641	2.17	-		2,354,159	5.97	866,683	2.20	-		1,181,532	2.99
40	1,615,128	4.09	1,813,599	4.60	2,143,898	5.43	2,198,652	5.57	2,699,452	6.84	1,347,647	3.42
41	2,038,577	5.17	2,270,982	5.76	2,411,670	6.11	2,039,291	5.17	1,828,512	4.63	410,454	1.04
42	1,554,560	3.94	2,728,648	6.92	2,158,437	5.47	2,025,958	5.14	1,785,398	4.53	959,173	2.43
43	1,739,381	4.41	2,083,584	5.28	1,808,257	4.58	934,714	2.37	2,119,671	5.37	864,251	2.19
44	1,493,053	3.78	1,307,472	3.31	2,079,809	5.27	1,850,731	4.69	3,187,041	8.08	1,720,509	4.36
45	1,875,787	4.75	2,827,515	7.17	1,475,404	3.74	2,138,905	5.42	2,675,637	6.78	1,196,369	3.03
46	1,981,406	5.02	1,728,435	4.38	2,037,838	5.17	2,014,658	5.11	1,513,374	3.84	1,320,990	3.35
47	2,824,474	7.16	1,839,565	4.66	2,739,365	6.94	1,670,833	4.24	2,626,572	6.66	1,854,929	4.70
48	928,301	2.35	2,452,214	6.22	1,720,299	4.36	1,461,811	3.71	1,474,019	3.74	1,765,907	4.48
49	2,211,104	5.60	3,019,519	7.65	2,665,681	6.76	3,807,742	9.65	1,666,484	4.22	1,859,303	4.71
50	2,102,789	5.33	2,139,614	5.42	2,841,671	7.20	2,385,626	6.05	1,601,717	4.06	1,805,553	4.58
51	1,724,999	4.37	3,313,718	8.40	2,153,745	5.46	4,062,633	10.30	1,798,325	4.56	1,400,553	3.55
52	2,230,557	5.65	2,500,067	6.34	2,202,918	5.58	532,472	1.35	979,853	2.48	805,251	2.04
53	1,429,353	3.62	1,635,336	4.15	1,634,261	4.14	371,828	0.94	3,435,295	8.71	1,376,160	3.49
54	2,175,965	5.52	2,823,953	7.16	1,889,435	4.79	1,574,522	3.99	2,096,065	5.31	923,492	2.34
55	436,316	1.11	2,092,844	5.30	2,103,425	5.33	563,320	1.43	1,367,057	3.47	966,996	2.45
56	711,216	1.80	923,892	2.34	1,958,105	4.96	576,518	1.46	1,215,216	3.08	519,941	1.32
57	2,456,839	6.23	850,126	2.15	1,691,901	4.29	1,890,656	4.79	1,093,237	2.77	902,554	2.29
58	3,480,250	8.82	2,301,849	5.83	1,862,376	4.72	1,595,791	4.04	1,897,601	4.81	1,086,446	2.75
59	2,803,107	7.11	2,078,836	5.27	2,789,231	7.07	2,512,160	6.37	975,242	2.47	1,352,058	3.43
60	3,089,484	7.83	3,469,562	8.79	2,097,312	5.32	3,193,694	8.10	1,766,730	4.48	1,105,573	2.80
61	1,798,074	4.56	1,840,006	4.66	2,047,564	5.19	804,648	2.04	3,054,768	7.74	1,067,469	2.71
62	2,830,235	7.17	1,979,556	5.02	1,751,467	4.44	946,947	2.40	3,985,599	10.10	1,745,096	4.42
63	3,239,359	8.21	1,356,852	3.44	1,313,069	3.33	1,614,906	4.09	853,672	2.16	944,441	2.39
64	1,586,110	4.02	952,957	2.42	2,019,480	5.12	619,048	1.57	1,886,150	4.78	1,402,107	3.55
65	2,396,862	6.08	1,501,126	3.80	1,661,313	4.21	1,808,637	4.58	1,923,107	4.87	946,898	2.40
66	1,120,727	2.84	2,982,535	7.56	3,227,734	8.18	1,990,244	5.04	1,474,266	3.74	1,537,817	3.90
67	1,211,548	3.07	1,563,352	3.96	2,084,222	5.28	1,341,432	3.40	1,370,373	3.47	1,474,196	3.74
68	1,743,080	4.42	2,260,925	5.73	1,842,843	4.67	1,662,156	4.21	2,079,111	5.27	948,119	2.40
69	2,079,443	5.27	1,271,801	3.22	2,281,570	5.78	1,019,397	2.58	1,284,895	3.26	1,098,356	2.78
70	1,649,461	4.18	1,718,407	4.36	1,408,067	3.57	1,376,860	3.49	1,934,097	4.90	880,630	2.23
71	1,070,878	2.71	436,543	1.11	1,020,183	2.59	1,276,863	3.24	2,691,102	6.82	773,059	1.96
72	986,297	2.50	1,517,337	3.85	2,391,420	6.06	1,779,716	4.51	3,104,920	7.87	748,339	1.90
73	1,330,483	3.37	2,507,376	6.36	2,038,644	5.17	2,248,360	5.70	1,365,657	3.46	1,065,599	2.70
74	1,540,538	3.90	3,486,279	8.84	980,800	2.49	1,056,556	2.68	1,581,903	4.01	1,076,958	2.73
75	1,684,553	4.27	4,587,876	11.63	1,142,872	2.90	1,714,415	4.35	2,528,119	6.41	1,208,772	3.06
76	2,267,015	5.75	2,598,535	6.59	1,875,281	4.75	2,594,134	6.58	1,271,255	3.22	983,786	2.49
77	1,522,836	3.86	1,368,889	3.47	1,522,223	3.86	1,839,649	4.66	1,111,223	2.82	1,608,117	4.08
78	734,815	1.86	1,404,873	3.56	2,280,967	5.78	775,498	1.97	2,562,654	6.50	814,268	2.06
79	1,156,985	2.93	3,007,068	7.62	1,604,919	4.07	1,216,701	3.08	2,323,477	5.89	459,784	1.17
80	2,655,047	6.73	1,851,900	4.69	3,071,982	7.79	-		2,473,673	6.27	843,113	2.14
81	880,141	2.23	827,086	2.10	1,792,139	4.54	3,525,796	8.94	1,950,440	4.94	614,167	1.56
82	2,235,513	5.67	3,141,795	7.96	2,443,911	6.19	2,543,489	6.45	2,037,071	5.16	985,294	2.50
83	1,557,578	3.95	2,707,453	6.86	2,586,765	6.56	2,652,867	6.72	2,210,898	5.60	1,533,889	3.89
84	2,462,518	6.24	1,374,929	3.49	1,221,090	3.10	2,612,582	6.62	1,703,687	4.32	837,118	2.12

85	-		1,067,392	2.71	1,966,152	4.98	1,745,259	4.42	1,670,924	4.24	1,124,584	2.85
86	773,700	1.96	1,890,743	4.79	2,326,831	5.90	1,870,680	4.74	1,085,357	2.75	787,283	2.00
87	1,637,187	4.15	1,560,839	3.96	1,688,664	4.28	2,162,170	5.48	1,745,164	4.42	833,851	2.11
88	1,890,718	4.79	2,376,476	6.02	2,168,139	5.50	1,791,286	4.54	2,868,732	7.27	884,413	2.24
89	1,246,893	3.16	-		705,391	1.79	1,353,633	3.43	749,357	1.90	945,148	2.40
90	1,309,372	3.32	223,954	0.57	1,547,188	3.92	1,371,069	3.48	1,910,440	4.84	496,343	1.26
91	2,051,785	5.20	400,200	1.01	2,213,359	5.61	1,710,812	4.34	3,054,270	7.74	1,651,609	4.19
92	2,160,768	5.48	194,651	0.49	1,001,064	2.54	1,553,226	3.94	1,764,221	4.47	1,747,011	4.43
93	1,397,081	3.54	162,372	0.41	849,855	2.15	2,120,252	5.37	2,302,104	5.84	654,845	1.66
94	1,359,283	3.45	254,619	0.65	1,390,082	3.52	764,514	1.94	2,345,489	5.95	791,867	2.01
95	1,051,394	2.66	174,601	0.44	558,155	1.41	462,107	1.17	1,551,722	3.93	1,069,993	2.71
96	1,333,056	3.38	293,454	0.74	2,188,796	5.55	1,695,123	4.30	3,741,854	9.48	1,827,465	4.63
Mean	**1,766,793**	**4.48**	**1,973,420**	**5.00**	**1,931,086**	**4.89**	**1,487,527**	**3.77**	**1,960,098**	**4.97**	**1,364,524**	**3.46**

Appendix 7.12 – Crossovers counts per chromosome in wild type and *msh2-1*, in the Col/Ler and Col/CLC F_2 populations. Crossovers were identified by genotyping by sequencing, and both total crossover (CO) number and CO number per F_2 individual are reported for each chromosome. The number of individuals per genotype ('*n*'), the mean CO number per F_2, and the standard deviation in total CO number per F_2 individual are stated. Statistical significance between wild type and *msh2-1* genotypes was assessed using a Wilcoxon rank sum test.

	Genotype							
	Col/Ler WT		**Col/Ler *msh2***		**Col/CLC WT**		**Col/CLC *msh2***	
Chromosome	**COs**	**COs/F_2**	**COs**	**COs/F_2**	**COs**	**COs/F_2**	**COs**	**COs/F_2**
1	434	1.81	362	1.94	405	2.14	428	2.24
2	320	1.33	270	1.44	231	1.22	210	1.10
3	252	1.05	191	1.02	212	1.12	245	1.28
4	298	1.24	214	1.14	256	1.35	244	1.28
5	414	1.73	357	1.91	391	2.07	407	2.13
Total	**1,718**	**7.16**	**1,394**	**7.45**	**1,495**	**7.91**	**1,534**	**8.03**
n	240		187		189		191	
Total CO StDev	2.22		2.09		2.28		2.54	
***P*-value (Wilcoxon)**	0.0534				0.602			

Appendix 7.13 – Quantification of heterochromatin in pachytene DAPI spreads, for wild type Col/Ler and Col/CLC. Heterochromatin was defined based on pixel intensity values for each image of DAPI-stained pachytene spreads, for Col/Ler and Col/CLC wild type genotypes. The nucleolus organiser regions were excluded from the analysis. Heterochromatin area was calculated as the percentage of each meiocyte's total area occupied by heterochromatin. Statistical significance was measured using a Wilcoxon rank sum test.

	Heterochromatin area (%)	
	Col/Ler WT	**Col/CLC WT**
	6.54	8.59
	6.63	8.67
	6.24	10.55
	5.51	7.86
	5.65	8.22
	6.33	8.86
	5.73	8.52
	6.59	8.51
	6.15	10.88
	6.70	9.15
Mean	**6.21**	**8.98**
StDev	**0.44**	**0.98**
***P*-value (Wilcoxon)**	**-**	**1.083×10^{-5}**

Appendix 7.14 – Comparison of SNP density surrounding wild type or *msh2-1* crossovers. SNP number was calculated around GBS crossover midpoints, using the corresponding 1,135 SNP datasets (Alonso-Blanco et al., 2016), using 1, 5, 10, 20, 40, 75 and 100 kb window sizes. SNP number was also calculated for equivalent random crossover datasets. Mean SNP number, standard deviation, and *P*-values are provided for all genotypes and comparisons. Statistical differences between means were assessed using the Dunn test, with the 'Benjamini–Hochberg' correction for multiple comparisons.

Width	Hybrid	Sample	Mean	StDev	Comparison	*P*-value
1kb	Col/Ler	WT	5.69	5.65	WT – *msh2*	2.13×10^{-10}
		random (WT)	4.39	6.71	*msh2* – random (*msh2*)	1.02×10^{-8}
		msh2	4.48	5.13	WT – random (*msh2*)	3.53×10^{-35}
		random (*msh2*)	4.32	6.30	*msh2* – random (WT)	2.04×10^{-11}
					WT – random (WT)	1.56×10^{-43}
					random (WT) – random (*msh2*)	0.451
	Col/CLC	WT	6.75	5.98	WT – *msh2*	1.66×10^{-5}
		random (WT)	4.93	6.57	*msh2* – random (*msh2*)	2.71×10^{-18}
		msh2	5.82	5.53	WT – random (*msh2*)	1.70×10^{-38}
		random (*msh2*)	4.97	6.50	*msh2* – random (WT)	2.55×10^{-19}
					WT – random (WT)	1.01×10^{-39}
					random (WT) – random (*msh2*)	0.726

Width	Hybrid	Sample	Mean	StDev	Comparison	*P*-value
5kb	Col/Ler	WT	26.5	20.0	WT – *msh2*	1.02×10^{-10}
		random (WT)	23.2	25.0	*msh2* – random (*msh2*)	8.78×10^{-3}
		msh2	22.2	19.0	WT – random (*msh2*)	1.86×10^{-20}
		random (*msh2*)	23.1	24.1	*msh2* – random (WT)	3.37×10^{-3}
					WT – random (WT)	1.65×10^{-23}
					random (WT) – random (*msh2*)	0.812
	Col/CLC	WT	31.6	21.1	WT – *msh2*	4.90×10^{-7}
		random (WT)	25.3	24.1	*msh2* – random (*msh2*)	1.14×10^{-10}
		msh2	27.7	20.0	WT – random (*msh2*)	3.35×10^{-30}
		random (*msh2*)	25.6	25.3	*msh2* – random (WT)	2.66×10^{-10}
					WT – random (WT)	1.73×10^{-29}
					random (WT) – random (*msh2*)	0.897

Width	Hybrid	Sample	Mean	StDev	Comparison	*P*-value
10kb	Col/Ler	WT	51.5	35.0	WT – *msh2*	1.28×10^{-9}
		random (WT)	46.3	43.7	*msh2* – random (*msh2*)	0.183
		msh2	44.1	33.3	WT – random (*msh2*)	4.66×10^{-14}
		random (*msh2*)	46.2	42.2	*msh2* – random (WT)	0.065
					WT – random (WT)	2.74×10^{-17}
					random (WT) – random (*msh2*)	0.605
	Col/CLC	WT	61.9	36.1	WT – *msh2*	1.47×10^{-7}
		random (WT)	50.5	43.2	*msh2* – random (*msh2*)	1.74×10^{-9}
		msh2	54.8	35.0	WT – random (*msh2*)	3.07×10^{-29}
		random (*msh2*)	50.0	42.8	*msh2* – random (WT)	1.62×10^{-8}
					WT – random (WT)	2.28×10^{-27}
					random (WT) – random (*msh2*)	0.708

Width	Hybrid	Sample	Mean	StDev	Comparison	*P*-value
20kb	Col/Ler	WT	101.4	60.6	WT – *msh2*	2.03×10^{-8}
		random (WT)	93.1	75.9	*msh2* – random (*msh2*)	0.384
		msh2	89.3	59.1	WT – random (*msh2*)	3.79×10^{-11}
		random (*msh2*)	92.1	73.3	*msh2* – random (WT)	0.435
					WT – random (WT)	4.53×10^{-12}
					random (WT) – random (*msh2*)	0.990
	Col/CLC	WT	121.4	62.6	WT – *msh2*	7.22×10^{-6}
		random (WT)	100.1	75.4	*msh2* – random (*msh2*)	1.69×10^{-10}
		msh2	110.7	62.0	WT – random (*msh2*)	2.99×10^{-27}
		random (*msh2*)	98.8	73.4	*msh2* – random (WT)	4.25×10^{-9}
					WT – random (WT)	7.43×10^{-25}
					random (WT) – random (*msh2*)	0.611

Width	Hybrid	Sample	Mean	StDev	Comparison	*P*-value
40kb	Col/Ler	WT	201.9	107.6	WT – *msh2*	2.17×10^{-8}
		random (WT)	185.1	128.1	*msh2* – random (*msh2*)	0.529
		msh2	179.1	103.3	WT – random (*msh2*)	1.31×10^{-10}
		random (*msh2*)	182.3	127.2	*msh2* – random (WT)	0.700
					WT – random (WT)	3.43×10^{-10}
					random (WT) – random (*msh2*)	0.665
	Col/CLC	WT	242.1	108.9	WT – *msh2*	2.36×10^{-6}
		random (WT)	197.9	131.6	*msh2* – random (*msh2*)	4.46×10^{-10}
		msh2	222.4	110.1	WT – random (*msh2*)	9.59×10^{-28}
		random (*msh2*)	198.1	125.7	*msh2* – random (WT)	4.00×10^{-10}
					WT – random (WT)	1.36×10^{-27}
					random (WT) – random (*msh2*)	0.919

Width	Hybrid	Sample	Mean	StDev	Comparison	*P*-value
75kb	Col/Ler	**WT**	376.7	178.7	**WT – *msh2***	3.81×10^{-8}
		random (WT)	346.4	206.8	***msh2* – random (*msh2*)**	0.640
		msh2	338.3	174.6	**WT – random (*msh2*)**	6.49×10^{-10}
		random (*msh2*)	339.0	202.3	***msh2* – random (WT)**	0.967
					WT – random (WT)	6.63×10^{-9}
					random (WT) – random (*msh2*)	0.513
	Col/CLC	**WT**	447.6	179.6	**WT – *msh2***	9.81×10^{-6}
		random (WT)	371.6	217.5	***msh2* – random (*msh2*)**	4.63×10^{-10}
		msh2	417.3	183.6	**WT – random (*msh2*)**	3.25×10^{-26}
		random (*msh2*)	372.6	211.2	***msh2* – random (WT)**	4.12×10^{-10}
					WT – random (WT)	9.88×10^{-26}
					random (WT) – random (*msh2*)	0.988

Width	Hybrid	Sample	Mean	StDev	Comparison	*P*-value
100kb	Col/Ler	**WT**	502.1	224.5	**WT – *msh2***	1.38×10^{-8}
		random (WT)	462.3	257.9	***msh2* – random (*msh2*)**	0.374
		msh2	451.8	218.5	**WT – random (*msh2*)**	4.26×10^{-11}
		random (*msh2*)	449.5	251.3	***msh2* – random (WT)**	0.917
					WT – random (WT)	5.52×10^{-9}
					random (WT) – random (*msh2*)	0.365
	Col/CLC	**WT**	595.9	227.1	**WT – *msh2***	1.16×10^{-5}
		random (WT)	496.7	275.2	***msh2* – random (*msh2*)**	6.37×10^{-11}
		msh2	557.0	231.1	**WT – random (*msh2*)**	1.86×10^{-27}
		random (*msh2*)	497.8	269.3	***msh2* – random (WT)**	8.84×10^{-11}
					WT – random (WT)	3.44×10^{-27}
					random (WT) – random (*msh2*)	0.961

Appendix 7.15 – Counts of seed set per silique for wild type, *msh4-1* and *HA-MSH4 msh4-1* complemented lines. Seeds were counted from 3 fruits, for 10 plants per genotype. Tests for statistical significance between the mean seed set for wild type – *msh4-1* and wild type – *HA-MSH4 msh4-1* were assessed using a Wilcoxon rank sum test.

Genotype	**Wild type**			***msh4-1***			***HA-MSH4 msh4-1***		
Silique #	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
1	59	60	62	2	1	3	63	63	64
2	61	61	64	3	2	2	64	62	65
3	64	58	61	4	4	5	60	64	65
4	65	60	62	3	3	2	66	60	64
5	68	59	64	3	1	4	71	59	60
6	60	59	60	2	3	2	65	60	59
7	63	62	59	5	3	2	60	64	65
8	64	63	63	0	0	2	61	63	64
9	62	63	64	3	4	1	64	64	64
10	64	61	60	3	5	0	62	60	60
Mean	**61.8**			**2.6**			**62.8**		
***P*-value**	**-**			**2.32×10^{-11}**			**0.099**		

Appendix 7.16 – Crossover frequency measured in the *420* genetic interval in wild type, *HA-MSH4 msh4-1* and *HA-MSH4 msh4-1 HEI10*. Genetic distance in centimorgans (cM) is calculated as $cM = 100\times(1-(1-2(N_G+N_R)/N_T)^{1/2})$, where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds of all classes analysed. Mann-Whitney *U* tests were performed to test for significant differences between genotypes.

Individual	Green alone	Red alone	Red and Green	Non-colour	Total seed	cM	cM/ Mb	Green: Non-green	Red: Non-red	Green alone: Red alone
Wild type	138	144	1,110	264	1,656	18.80	3.68	3.06	3.12	0.96
Wild type	208	204	1,599	385	2,396	19.00	3.72	3.07	3.04	1.02
Wild type	209	230	1,515	382	2,336	21.00	4.11	2.82	2.95	0.91
Wild type	226	230	1,660	461	2,577	19.62	3.84	2.73	2.75	0.98
Wild type	201	185	1,374	361	2,121	20.25	3.96	2.88	2.77	1.09
Wild type	184	197	1,449	371	2,201	19.14	3.75	2.88	2.97	0.93
Wild type	214	219	1,617	358	2,408	19.98	3.91	3.17	3.21	0.98
Wild type	244	212	1,752	458	2,666	18.89	3.70	2.98	2.80	1.15
Wild type	231	213	1,543	346	2,333	21.30	4.17	3.17	3.04	1.08
Wild type	203	202	1,605	392	2,402	18.59	3.64	3.04	3.04	1.00
Total	**2058**	**2036**	**15224**	**3778**	**23,096**	**19.66**	**3.85**	**-**		
HA-MSH4 msh4	195	177	1,512	395	2,279	17.93	3.51	2.98	2.86	1.10
HA-MSH4 msh4	172	185	1,509	427	2,293	17.02	3.33	2.75	2.83	0.93
HA-MSH4 msh4	174	171	1,557	391	2,293	16.39	3.21	3.08	3.06	1.02
HA-MSH4 msh4	193	190	1,520	390	2,293	18.39	3.60	2.95	2.93	1.02
HA-MSH4 msh4	191	178	1,530	394	2,293	17.65	3.45	3.01	2.92	1.07
HA-MSH4 msh4	206	195	1,549	343	2,293	19.36	3.79	3.26	3.18	1.06
HA-MSH4 msh4	209	186	1,525	412	2,332	18.68	3.66	2.90	2.76	1.12
HA-MSH4 msh4	204	218	1,562	348	2,332	20.12	3.94	3.12	3.22	0.94
HA-MSH4 msh4	195	216	1,532	389	2,332	19.53	3.82	2.85	2.99	0.90
HA-MSH4 msh4	224	221	1,528	359	2,332	21.36	4.18	3.02	3.00	1.01
HA-MSH4 msh4	213	230	1,533	356	2,332	21.26	4.16	2.98	3.10	0.93
Total	**2,176**	**2,167**	**16,857**	**4,204**	**25,404**	**18.88**	**3.70**	***P* = 0.223**		
HEI10 HA-MSH4 msh4	328	324	1,403	277	2,332	33.61	6.58	2.88	2.85	1.01
HEI10 HA-MSH4 msh4	255	334	1,453	290	2,332	29.65	5.80	2.74	3.28	0.76
HEI10 HA-MSH4 msh4	255	287	1,524	266	2,332	26.85	5.25	3.22	3.48	0.89
HEI10 HA-MSH4 msh4	347	324	1,397	264	2,332	34.84	6.82	2.97	2.82	1.07
HEI10 HA-MSH4 msh4	488	475	1,266	103	2,332	41.73	8.17	3.03	2.95	1.03
Total	**1,673**	**1,744**	**7,043**	**1,200**	**11,660**	**33.34**	**6.52**	***P* = 0.0004**		

Appendix 7.17 – Flow cytometry fluorescent pollen count data for wild type and *HA-MSH4 msh4-1* at the *I2f* and *I3c* genetic intervals. Genetic distance in centimorgans (cM) is calculated as cM = $100\times(N_Y/(N_Y+N_{R+Y}))$, where N_Y is the number of yellow alone pollen, and N_{Y+R} is the number of red and yellow pollen. Mann-Whitney *U* tests were performed to test for significant differences between genotypes.

Individual	Total	Red alone	Red + Yellow	Non-colour	Yellow alone	cM	cM/Mb	*P*-value
I2f Col-0 WT	25,472	1,114	9,132	14,568	658	6.72	10.03	
I2f Col-0 WT	24,699	1,139	9,358	13,505	697	6.93	10.35	
I2f Col-0 WT	26,119	1,293	9,488	14,652	686	6.74	10.06	
Total	**76,290**	**3,546**	**27,978**	**42,725**	**2,041**	**6.80**	**10.15**	**-**
I2f HA-MSH4 msh4	26,348	1,134	9,983	14,605	626	5.90	8.81	
I2f HA-MSH4 msh4	25,202	1,053	9,540	13,914	695	6.79	10.13	
I2f HA-MSH4 msh4	27,060	1,344	9,795	15,210	711	6.77	10.10	
I2f HA-MSH4 msh4	25,373	1,011	9,785	13,874	703	6.70	10.00	
Total	**103,983**	**4,542**	**39,103**	**57,603**	**2,735**	**6.54**	**9.76**	**0.629**
I3c Col-0 WT	14,663	951	5,235	8,178	299	5.40	4.54	
I3c Col-0 WT	-	-	1,837	-	124	6.32	5.31	
I3c Col-0 WT	-	-	5,441	-	354	6.11	5.13	
I3c Col-0 WT	-	-	3,594	-	200	5.27	4.43	
Total	**-**	**-**	**16,107**	**-**	**997**	**5.78**	**4.85**	**-**
I3c HA-MSH4 msh4	23,883	1,697	8,287	13,350	549	6.21	5.22	
I3c HA-MSH4 msh4	23,287	1,539	7,819	13,441	488	5.87	4.94	
I3c HA-MSH4 msh4	39,729	2,352	14,715	21,675	987	6.29	5.28	
I3c HA-MSH4 msh4	16,337	1,264	5,206	9,537	330	5.96	5.01	
Total	**103,236**	**6,852**	**36,027**	**58,003**	**2,354**	**6.08**	**5.11**	**0.686**

Appendix 7.18 – Summary of total and aligned read pairs (2 × 76 bp) in the HA-MSH4 (biological replicate 1) ChIP and input sequencing libraries. Deduplicated read pairs were aligned to the TAIR10 reference genome using Bowtie2. Uniquely and multiply aligning reads with more than 2 mismatches and multiply aligning reads with MAPQ scores less than 10 were discarded.

Library	Total sequenced read pairs	Deduplicated (% of total sequenced read pairs)	Aligned (% of deduplicated reads)	Uniquely aligning, mismatch filtered (removed reads with mismatches>2) (% of aligned reads in column 4)	Multiply aligning, mismatch filtered (removed reads with mismatches>2 and MAPQ<10) (% of aligned reads in column 4)	Both (% of aligned reads in column 4)	Average depth of coverage [(filtered aligned read pairs*(2*76)) / genome size]
MSH4 Rep1 ChIP	71,167,988	25,601,969 (35.97%)	23,330,683 (91.13%)	13,044,234 (55.91%)	1,989,353 (8.53%)	15,033,587 (64.44%)	16.93×
MSH4 Rep1 input	72,113,988	15,629,343 (21.67%)	13,813,845 (88.38%)	7,571,569 (54.81%)	982,629 (7.11%)	8,554,198 (61.92%)	9.63×

Bibliography

Acharya, S., Foster, P. L., Brooks, P., & Fishel, R. (2003). The Coordinated Functions of the *E. coli* MutS and MutL Proteins in Mismatch Repair. *Molecular Cell*, *12*(1), 233–246.

Acquaviva, L., Székvölgyi, L., Dichtl, B., Dichtl, B. S., André, C. de L. R. Saint, Nicolas, A., & Géli, V. (2013). The COMPASS Subunit Spp1 Links Histone Methylation to Initiation of Meiotic Recombination. *Science*, *339*(6116), 215–218.

Adé, J., Belzile, F., Philippe, H., & Doutriaux, M.-P. (1999). Four mismatch repair paralogues coexist in *Arabidopsis thaliana*: AtMSH2, AtMSH3, AtMSH6-1 and AtMSH6-2. *MGG - Molecular and General Genetics*, *262*(2), 239–249.

Adé, J., Haffani, Y., & Belzile, F. J. (2001). Functional analysis of the *Arabidopsis thaliana* mismatch repair gene *MSH2*. *Genome*, *44*(4), 651–657.

Agarwal, S., & Roeder, G. S. (2000). Zip3 Provides a Link between Recombination Enzymes and Synaptonemal Complex Proteins. *Cell*, *102*(2), 245–255.

Ahuja, J. S., Sandhu, R., Mainpal, R., Lawson, C., Henley, H., Hunt, P. A., Yanowitz, J. L., & Börner, G. V. (2017). Control of meiotic pairing and recombination by chromosomally tethered 26S proteasome. *Science*, *355*(6323), 408–411.

Aklilu, B. B., Soderquist, R. S., & Culligan, K. M. (2014). Genetic analysis of the Replication Protein A large subunit family in Arabidopsis reveals unique and overlapping roles in DNA repair, meiosis and DNA replication. *Nucleic Acids Research*, *42*(5), 3104–3118.

Alani, E., Lee, S., Kane, M. F., Griffith, J., & Kolodner, R. D. (1997). *Saccharomyces cerevisiae* MSH2, a mispaired base recognition protein, also recognizes Holliday junctions in DNA. *Journal of Molecular Biology*, *265*(3), 289–301.

Alexander, M. P. (1969). Differential Staining of Aborted and Nonaborted Pollen. *Stain Technology*, *44*(3), 117–122.

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C. C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D. E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W. L., Berry, C. C., & Ecker, J. R. (2003). Genome-Wide Insertional Mutagenesis of *Arabidopsis thaliana*. *Science*, *301*(5633), 653–657.

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., Cao, J., Chae, E., Dezwaan, T. M., Ding, W., Ecker, J.R., Exposito-Alonso, M., Farlow, A., Fitz, J., Gan, X., Grimm, D.G., Hancock, A.M., Henz, S.R., Holm, S., Horton, M., Jarsulic, M., Kerstetter, R.A., Korte, A., Korte, P., Lanz, C., Lee, C.R., Meng, D., Michael, T. P., Mott, R., Muliyati, N. W., Nägele, T., Nagler, M., Nizhynska, V., Nordborg, M., Novikova, P. Y., Picó, F. X., Platzer, A., Rabanal, F.A., Rodriguez, A., Rowan, B. A., Salomé, P.A., Schmid, K.J., Schmitz, R. J., Seren, Ü., Sperone, F.G., Sudkamp, M., Svardal, H., Tanzer, M. M., Todd, D., Volchenboum, S. L., Wang, C., Wang, G., Wang, X., Weckwerth, W., Weigel, D., & Zhou, X. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. *Cell*, *166*(2), 481–491.

Armstrong, S. J., Caryl, A. P., Jones, G. H., & Franklin, F. C. H. (2002). Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in Arabidopsis and Brassica. *Journal of Cell Science*, *115*(Pt 18), 3645–3655.

Armstrong, S. J., & Jones, G. H. (2003). Meiotic cytology and chromosome behaviour in wild-type *Arabidopsis thaliana*. *Journal of Experimental Botany*, *54*(380), 1–10.

Armstrong, S. (2013). A Time Course for the Analysis of Meiotic Progression in *Arabidopsis thaliana*. In A. S. (eds) Pawlowski W., Grelon M. (Ed.) (Vol. 990, pp. 119–123). Humana Press, Totowa, NJ.

Arora, K., & Corbett, K. D. (2018). The conserved XPF:ERCC1-like Zip2:Spo16 complex controls meiotic crossover formation through structure-specific DNA binding. *Nucleic Acids Research*.

Auton, A., Rui Li, Y., Kidd, J., Oliveira, K., Nadel, J., Holloway, J. K., Hayward, J. J., Cohen, P. E., Greally, J. M., Wang, J., Bustamante, C. D., & Boyko, A. R. (2013). Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs. *PLoS Genetics*, *9*(12), e1003984.

Azumi, Y., Liu, D., Zhao, D., Li, W., Wang, G., Hu, Y., & Ma, H. (2007). Homolog interaction during meiotic prophase I in Arabidopsis requires the SOLO DANCERS gene encoding a novel cyclin-like protein. *The EMBO Journal*, *21*(12), 3081–3095.

Bailey, T., Krajewski, P., Ladunga, I., Lefebvre, C., Li, Q., Liu, T., Madrigal, P., Taslim, C., & Zhang, J. (2013). Practical Guidelines for the Comprehensive Analysis of ChIP-seq Data. *PLoS Comput Biol*, *9*(11): e1003326.

Barth, S., Melchinger, A. E., Devezi-Savula, B., & Lübberstedt, T. (2001). Influence of genetic background and heterozygosity on meiotic recombination in *Arabidopsis thaliana*. *Genome*, *44*(6), 971–978.

Baudat, F., Manova, K., Yuen, J. P., Jasin, M., & Keeney, S. (2000). Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Molecular Cell*, *6*(5), 989–998.

Baudat, F., & de Massy, B. (2007). Cis- and Trans-Acting Elements Regulate the Mouse Psmb9 Meiotic Recombination Hotspot. *PLoS Genetics*, *3*(6), e100.

Baudat, F., Imai, Y., & de Massy, B. (2013). Meiotic recombination in mammals: localization and regulation. *Nature Reviews Genetics*, *14*(11), 794–806.

Bauer, E., Falque, M., Walter, H., Bauland, C., Camisan, C., Campo, L., Meyer, N., Ranc, N., Rincent, R., Schipprack, W., Altmann, T., Flament, P., Melchinger, A. E., Menz, M., Moreno-González, J., Ouzunova, M., Revilla, P., Charcosset, A., Martin, O. C., & Schön, C.-C. (2013). Intraspecific variation of recombination rate in maize. *Genome Biology*, *14*(9), R103.

Belfield, E. J., Ding, Z. J., Jamieson, F. J. C., Visscher, A. M., Zheng, S. J., Mithani, A., & Harberd, N. P. (2018). DNA mismatch repair preferentially protects genes from mutation. *Genome Research*, *28*(1), 66–74.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573–580.

Berchowitz, L. E., Francis, K. E., Bey, A. L., & Copenhaver, G. P. (2007). The Role of AtMUS81 in Interference-Insensitive Crossovers in *A. thaliana*. *PLoS Genetics*, *3*(8), e132.

Berchowitz, L. E., & Copenhaver, G. P. (2008). Fluorescent Arabidopsis tetrads: a visual assay for quickly developing large crossover and crossover interference data sets. *Nature Protocols*, *3*(1), 41–50.

Bernard, P., Maure, J.-F., Partridge, J. F., Genier, S., Javerzat, J.-P., & Allshire, R. C. (2001). Requirement of Heterochromatin for Cohesion at Centromeres. *Science*, *294*(5551), 2539–2542.

Bewick, A. J., & Schmitz, R. J. (2017). Gene body DNA methylation in plants. *Current Opinion in Plant Biology*, *36*, 103–110.

Bishop, D. K., Park, D., Xu, L., & Kleckner, N. (1992). DMC1: a meiosis-specific yeast homolog of *E. coli* recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell*, *69*(3), 439–456.

Biswas, I., & Hsieh, P. (1997). Interaction of MutS protein with the major and minor grooves of a heteroduplex DNA. *The Journal of Biological Chemistry*, *272*(20), 13355–13364.

Boer, E. de, Stam, P., Dietrich, A. J. J., Pastink, A., & Heyting, C. (2006). Two levels of interference in mouse meiotic recombination. *Proceedings of the National Academy of Sciences*, *103*(25), 9607–9612.

Borde, V., Robine, N., Lin, W., Bonfils, S., Géli, V., & Nicolas, A. (2009). Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *The EMBO Journal*, *28*(2), 99–111.

Borevitz, J. O., Hazen, S. P., Michael, T. P., Morris, G. P., Baxter, I. R., Hu, T. T., Chen, H., Werner, J. D., Nordborg, M., Salt, D. E., Kay, S. A., Chory, J., Weigel, D., Jones, J. D. G., & Ecker, J. R. (2007). Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, *104*(29), 12057–12062.

Börner, G. V., Kleckner, N., & Hunter, N. (2004). Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell*, *117*(1), 29–45.

Borts, R. H., & Haber, J. E. (1987). Meiotic recombination in yeast: alteration by multiple heterozygosities. *Science (New York, N.Y.)*, *237*(4821), 1459–1465.

Borts, R. H., Leung, W. Y., Kramer, W., Kramer, B., Williamson, M., Fogel, S., & Haber, J. E. (1990). Mismatch repair-induced meiotic recombination requires the pms1 gene product. *Genetics*, *124*(3).

Bruni, R., Martin, D., & Jiricny, J. (1988). d(GATC) sequences influence *Escherichia coli* mismatch repair in a distance-dependent manner from positions both upstream and downstream of the mismatch. *Nucleic Acids Research*, *16*(11), 4875–4890.

Bugreev, D. V, Pezza, R. J., Mazina, O. M., Voloshin, O. N., Camerini-Otero, R. D., & Mazin, A. V. (2011). The resistance of DMC1 D-loops to dissociation may account for the DMC1 requirement in meiosis. *Nature Structural & Molecular Biology*, *18*(1), 56–60.

Bühler, M., & Gasser, S. M. (2009). Silent chromatin at the middle and ends: lessons from yeasts. *The EMBO Journal*, *28*(15), 2149–2161.

Busslinger, G. A., Stocsits, R. R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., & Peters, J.-M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*, *544*(7651), 503–507.

Cai, X., Dong, F., Edelmann, R. E., & Makaroff, C. A. (2003). The Arabidopsis SYN1 cohesin protein is required for sister chromatid arm cohesion and homologous chromosome pairing. *Journal of Cell Science*, *116*(Pt 14), 2999–3007.

Calmann, M. A., Nowosielska, A., & Marinus, M. G. (2005). Separation of mutation avoidance and antirecombination functions in an *Escherichia coli* mutS mutant. *Nucleic Acids Research*, *33*(4), 1193–1200.

Cannavo, E., Johnson, D., Andres, S. N., Kissling, V. M., Reinert, J. K., Garcia, V., Erie, D. A., Hess, D., Thoma, N. H., Enchev, R. I., Peter, M., Williams, R. S., Neale, M. J., & Cejka, P. (2018). Regulatory control of DNA end resection by Sae2 phosphorylation. *Nature Communications*, *9*(1), 4016.

Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., Wang, X., Ott, F., Muller, J., Alonso-Blanco, C., Borgwardt, K., Schmid, K. J., & Weigel, D. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics*, *43*(10), 956–963.

Carballo, J. A., Johnson, A. L., Sedgwick, S. G., & Cha, R. S. (2008). Phosphorylation of the Axial Element Protein Hop1 by Mec1/Tel1 Ensures Meiotic Interhomolog Recombination. *Cell*, *132*(5), 758–770.

Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., Moffat, J., Golland, P., & Sabatini, D. M. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, *7*(10), R100.

Carr, M., Bensasson, D., & Bergman, C. M. (2012). Evolutionary Genomics of Transposable Elements in *Saccharomyces cerevisiae*. *PLoS ONE*, *7*(11), e50978.

Chakraborty, U., & Alani, E. (2016). Understanding how mismatch repair proteins participate in the repair/anti-recombination decision. *FEMS Yeast Research*, *16*(6), fow071.

Chambon, A., West, A., Vezon, D., Horlow, C., De Muyt, A., Chelysheva, L., Ronceret, A., Darbyshire, A., Osman, K., Heckmann, S., Franklin, F. C. H., & Grelon, M. (2018). Identification of ASYNAPTIC4, a Component of the Meiotic Chromosome Axis. *Plant Physiology*, *178*(1), 233–246.

Chelysheva, L., Gendrot, G., Vezon, D., Doutriaux, M.-P., Mercier, R., & Grelon, M. (2007). Zip4/Spo22 Is Required for Class I CO Formation but Not for Synapsis Completion in *Arabidopsis thaliana*. *PLoS Genetics*, *3*(5), e83.

Chelysheva, L., Vezon, D., Chambon, A., Gendrot, G., Pereira, L., Lemhemdi, A., Vrielynck, N., Le Guin, S., Novatchkova, M., & Grelon, M. (2012). The Arabidopsis HEI10 Is a New ZMM Protein Related to Zip3. *PLoS Genetics*, *8*(7), e1002799.

Chen, W., & Jinks-Robertson, S. (1999). The Role of the Mismatch Repair Machinery in Regulating Mitotic and Meiotic Recombination Between Diverged Sequences in Yeast. *Genetics*, *151*(4).

Chen, C., Zhang, W., Timofejeva, L., Gerardin, Y., & Ma, H. (2005). The Arabidopsis ROCK-N-ROLLERS gene encodes a homolog of the yeast ATP-dependent DNA helicase MER3 and is required for normal meiotic crossover formation. *The Plant Journal*, *43*(3), 321–334.

Cheng, C.-H., Lo, Y.-H., Liang, S.-S., Ti, S.-C., Lin, F.-M., Yeh, C.-H., Huang, H. Y., & Wang, T.-F. (2006). SUMO modifications control assembly of synaptonemal complex and polycomplex in meiosis of *Saccharomyces cerevisiae*. *Genes & Development*, *20*(15), 2067–2081.

Chen, S. Y., Tsubouchi, T., Rockmill, B., Sandler, J. S., Richards, D. R., Vader, G., Hochwagen, A., Roeder, G. S., & Fung, J. C. (2008). Global analysis of the meiotic crossover landscape. *Developmental Cell*, *15*(3), 401–415.

Chen, X., Suhandynata, R. T., Sandhu, R., Rockmill, B., Mohibullah, N., Niu, H., Liang, J., Lo, H.-C., Miller, D. E., Zhou, H., Börner, V., & Hollingsworth, N. M. (2015). Phosphorylation of the Synaptonemal Complex Protein Zip1 Regulates the Crossover/Noncrossover Decision during Yeast Meiosis. *PLOS Biology*, *13*(12), e1002329.

Chen, C., Lim, H. H., Shi, J., Tamura, S., Maeshima, K., Surana, U., & Gan, L. (2016). Budding yeast chromatin is dispersed in a crowded nucleoplasm in vivo. *Molecular Biology of the Cell*, *27*(21), 3357–3368.

Cherry, J. M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., Dunn, B., Dwight, S., Riles, L., Mortimer, R. K., & Botstein, D. (1997). Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature*, *387*(6632 Suppl), 67–73.

Chodavarapu, R. K., Feng, S., Bernatavichute, Y. V., Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S. J., Casero, D., Bernal, M., Huijser, P., Clark, A. T., Krämer, U., Merchant, S. S., Zhang, X., Jacobsen, S. E., & Pellegrini, M. (2010). Relationship between nucleosome positioning and DNA methylation. *Nature*, *466*(7304), 388–392.

Choi, K., Zhao, X., Kelly, K. A., Venn, O., Higgins, J. D., Yelina, N. E., Hardcastle, T. J., Ziolkowski, P. A., Copenhaver, G. P., Franklin, F. C., McVean, G., & Henderson, I. R. (2013). Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nature Genetics*, *45*(11), 1327–1336.

Choi, K., Reinhard, C., Serra, H., Ziolkowski, P. A., Underwood, C. J., Zhao, X., Hardcastle, T. J., Yelina, N. E., Griffin, C., Jackson, M., Mézard, C., McVean, G., Copenhaver, G. P., & Henderson, I. R. (2016). Recombination Rate Heterogeneity within Arabidopsis Disease Resistance Genes. *PLOS Genetics*, *12*(7), e1006179.

Choi, K., Yelina, N. E., Serra, H., & Henderson, I. R. (2017). Quantification and Sequencing of Crossover Recombinant Molecules from Arabidopsis Pollen DNA (pp. 23–57). Humana Press, New York, NY.

Choi, K., Zhao, X., Tock, A. J., Lambing, C., Underwood, C. J., Hardcastle, T. J., Serra, H., Kim, J., Cho, H. S., Kim, J., Ziolkowski, P. A., Yelina, N. E., Hwang, I., Martienssen, R. A., & Henderson, I. R. (2018). Nucleosomes and DNA methylation shape meiotic DSB frequency in *Arabidopsis thaliana* transposons and gene regulatory regions. *Genome Research*, *28*(4), 532–546.

Clark, A. B., Deterding, L., Tomer, K. B., & Kunkel, T. A. (2007). Multiple functions for the N-terminal region of Msh6. *Nucleic Acids Research*, *35*(12), 4114–4123.

Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal : For Cell and Molecular Biology*, *16*(6), 735–743.

Cole, F., Keeney, S., & Jasin, M. (2010). Comprehensive, Fine-Scale Dissection of Homologous Recombination Outcomes at a Hot Spot in Mouse Meiosis. *Molecular Cell*, *39*(5), 700–710.

Cole, F., Kauppi, L., Lange, J., Roig, I., Wang, R., Keeney, S., & Jasin, M. (2012). Homeostatic control of recombination is implemented progressively in mouse meiosis. *Nature Cell Biology*, *14*(4), 424–430.

Collins, I., & Newlon, C. S. (1994). Meiosis-specific formation of joint DNA molecules containing sequences from homologous chromosomes. *Cell*, *76*(1), 65–75.

Comeron, J. M., Ratnappan, R., & Bailin, S. (2012). The Many Landscapes of Recombination in *Drosophila melanogaster*. *PLoS Genetics*, *8*(10), e1002905.

Cooper, T. J., Garcia, V., & Neale, M. J. (2016). Meiotic DSB patterning: A multifaceted process. *Cell Cycle*, *15*(1), 13–21.

Cooper, T. J., Crawford, M. R., Hunt, L. J., Marsolier-Kergoat, M.-C., Llorente, B., & Neale, M. J. (2018). Mismatch repair impedes meiotic crossover interference. *BioRxiv*, 480418.

Copenhaver, G. P., Browne, W. E., & Preuss, D. (1998). Assaying genome-wide recombination and centromere functions with Arabidopsis tetrads. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(1), 247–252.

Copenhaver, G. P., Nickel, K., Kuromori, T., Benito, M. I., Kaul, S., Lin, X., Bevan, M., Murphy, G., Harris, B., Parnell, L. D., McCombie, W. R., Martienssen, R. A., Marra, M., & Preuss, D. (1999). Genetic definition and sequence analysis of Arabidopsis centromeres. *Science (New York, N.Y.)*, *286*(5449), 2468–2474.

Copenhaver, G. P., Housworth, E. A., & Stahl, F. W. (2002). Crossover interference in Arabidopsis. *Genetics*, *160*(4), 1631–1639.

Couteau, F., Belzile, F., Horlow, C., Grandjean, O., Vezon, D., & Doutriaux, M. P. (1999). Random chromosome segregation without meiotic arrest in both male and female meiocytes of a dmc1 mutant of Arabidopsis. *The Plant Cell*, *11*(9), 1623–1634.

Crismani, W., Girard, C., Froger, N., Pradillo, M., Santos, J. L., Chelysheva, L., Copenhaver, G. P., Horlow, C., & Mercier, R. (2012). FANCM Limits Meiotic Crossovers. *Science*, *336*(6088), 1588–1590.

Crown, K. N., Miller, D. E., Sekelsky, J., & Hawley, R. S. (2018). Local Inversion Heterozygosity Alters Recombination throughout the Genome. *Current Biology*, *28*(18), 2984–2990.e3.

Culligan, K. M., & Hays, J. B. (2000). Arabidopsis MutS homologs-AtMSH2, AtMSH3, AtMSH6, and a novel AtMSH7-form three distinct protein heterodimers with different specificities for mismatched DNA. *The Plant Cell*, *12*(6), 991–1002.

Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, *14*(4), 262–274.

Da Ines, O., Degroote, F., Goubely, C., Amiard, S., Gallego, M. E., & White, C. I. (2013). Meiotic Recombination in Arabidopsis Is Catalysed by DMC1, with RAD51 Playing a Supporting Role. *PLoS Genetics*, *9*(9), e1003787.

DasGupta, C., & Radding, C. M. (1982). Polar branch migration promoted by recA protein: effect of mismatched base pairs. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(3), 762–766.

Datta, A., Hendrix, M., Lipsitch, M., & Jinks-Robertson, S. (1997). Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(18), 9757–9762.

De Muyt, A., Pereira, L., Vezon, D., Chelysheva, L., Gendrot, G., Chambon, A., Lainé-Choinard, S., Pelletier, G., Mercier, R., Nogué, F., & Grelon, M. (2009). A High Throughput Genetic Screen Identifies New Early Meiotic Recombination Functions in *Arabidopsis thaliana*. *PLoS Genetics*, *5*(9), e1000654.

De Muyt, A., Pyatnitskaya, A., Andréani, J., Ranjha, L., Ramus, C., Laureau, R., Fernandez-Vega, A., Holoch, D., Girard, E., Govin, J., Margueron, R., Couté, Y., Cejka, P., Guérois, R., & Borde, V. (2018). A meiotic XPF-ERCC1-like complex recognizes joint molecule recombination intermediates to promote crossover formation. *Genes & Development*, *32*(3–4), 283–296.

de Wind, N., Dekker, M., Berns, A., Radman, M., & te Riele, H. (1995). Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell*, *82*(2), 321–330.

Demirci, S., van Dijk, A. D. J., Sanchez Perez, G., Aflitos, S. A., de Ridder, D., & Peters, S. A. (2017). Distribution, position and genomic characteristics of crossovers in tomato recombinant inbred lines derived from an interspecific cross between *Solanum lycopersicum* and *Solanum pimpinellifolium*. *The Plant Journal*, *89*(3), 554–564.

Deniz, O., Flores, O., Battistini, F., Pérez, A., Soler-López, M., & Orozco, M. (2011). Physical properties of naked DNA influence nucleosome positioning and correlate with transcription start and termination sites in yeast. *BMC Genomics*, *12*, 489.

Dernburg, A. F., McDonald, K., Moulder, G., Barstead, R., Dresser, M., & Villeneuve, A. M. (1998). Meiotic recombination in *C. elegans* initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Cell*, *94*(3), 387–398.

Diagouraga, B., Clément, J. A. J., Duret, L., Kadlec, J., de Massy, B., & Baudat, F. (2018). PRDM9 Methyltransferase Activity Is Essential for Meiotic DNA Double-Strand Break Formation at Its Binding Sites. *Molecular Cell*, *69*(5), 853–865.e6.

Dion, É., Li, L., Jean, M., & Belzile, F. (2007). An Arabidopsis MLH1 mutant exhibits reproductive defects and reveals a dual role for this gene in mitotic recombination. *The Plant Journal*, *51*(3), 431–440.

Dooner, H. K. (1986). Genetic fine structure of the bronze locus in maize. *Genetics*, *113*(4).

Dooner, H. K., & Martínez-Férez, I. M. (1997). Recombination occurs uniformly within the bronze gene, a meiotic recombination hotspot in the maize genome. *The Plant Cell*, *9*(9), 1633–1646.

Dooner, H. K. (2002). Extensive interallelic polymorphisms drive meiotic recombination into a crossover pathway. *The Plant Cell*, *14*(5), 1173–1183.

Drouaud, J., Camilleri, C., Bourguignon, P.-Y., Canaguier, A., Bérard, A., Vezon, D., Giancola, S., Brunel, D., Colot, V., Prum, B., Quesneville, H., & Mézard, C. (2006). Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination “hot spots”. *Genome Research*, *16*(1), 106–114.

Drouaud, J., Mercier, R., Chelysheva, L., Bérard, A., Falque, M., Martin, O., Zanni, V., Brunel, D., & Mézard, C. (2007). Sex-Specific Crossover Distributions and Variations in Interference Level along *Arabidopsis thaliana* Chromosome 4. *PLoS Genetics*, *3*(6), e106.

Drouaud, J., & Mézard, C. (2011). Characterization of Meiotic Crossovers in Pollen from *Arabidopsis thaliana*. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 745, pp. 223–249). Humana Press.

Drouaud, J., Khademian, H., Giraut, L., Zanni, V., Bellalou, S., Henderson, I. R., Falque, M., & Mézard, C. (2013). Contrasted Patterns of Crossover and Non-crossover at *Arabidopsis thaliana* Meiotic Recombination Hotspots. *PLoS Genetics*, *9*(11), e1003922.

Drummond, J. T., Li, G. M., Longley, M. J., & Modrich, P. (1995). Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science (New York, N.Y.)*, *268*(5219), 1909–1912.

Duina, A. A., Miller, M. E., & Keeney, J. B. (2014). Budding yeast for budding geneticists: a primer on the *Saccharomyces cerevisiae* model system. *Genetics*, *197*(1), 33–48.

Duroc, Y., Kumar, R., Ranjha, L., Adam, C., Guérois, R., Md Muntaz, K., Marsolier-Kergoat, M. C., Dingli, F., Laureau, R., Loew, D., Llorente, B., Charbonnier, J. B., Cejka, P., & Borde, V. (2017). Concerted action of the MutLβ heterodimer and Mer3 helicase regulates the global extent of meiotic gene conversion. *ELife*, *6*.

Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., Tsuchimatsu, T., Burbano, H. A., Picó, F. X., Alonso-Blanco, C., & Hancock, A. M. (2017). African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(20), 5213–5218.

Edwards, K., Johnstone, C., & Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, *19*(6), 1349.

Ellermeier, C., Higuchi, E. C., Phadnis, N., Holm, L., Geelhood, J. L., Thon, G., & Smith, G. R. (2010). RNAi and heterochromatin repress centromeric meiotic recombination. *Proceedings of the National Academy of Sciences*, *107*(19), 8701–8705.

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A., & Jasin, M. (1998). Gene conversion tracts from double-strand break repair in mammalian cells. *Molecular and Cellular Biology*, *18*(1), 93–101.

Elliott, B., & Jasin, M. (2001). Repair of Double-Strand Breaks by Homologous Recombination in Mismatch Repair-Defective Mammalian Cells. *Molecular and Cellular Biology*, *21*(8), 2671–2682.

Emmanuel, E., Yehuda, E., Melamed-Bessudo, C., Avivi-Ragolsky, N., & Levy, A. A. (2006). The role of AtMSH2 in homologous recombination in *Arabidopsis thaliana*. *EMBO Reports*, *7*(1), 100–105.

Evans, E., Sugawara, N., Haber, J. E., & Alani, E. (2000). The *Saccharomyces cerevisiae* Msh2 Mismatch Repair Protein Localizes to Recombination Intermediates In Vivo. *Molecular Cell*, *5*(5), 789–799.

Fang, W. H., & Modrich, P. (1993). Human strand-specific mismatch repair occurs by a bidirectional mechanism similar to that of the bacterial reaction. *The Journal of Biological Chemistry*, *268*(16), 11838–11844.

Fasching, C. L., Cejka, P., Kowalczykowski, S. C., & Heyer, W.-D. (2015). Top3-Rmi1 Dissolve Rad51-Mediated D Loops by a Topoisomerase-Based Mechanism. *Molecular Cell*, *57*(4), 595–606.

Feinstein, S. I., & Low, K. B. (1986). Hyper-recombining recipient strains in bacterial conjugation. *Genetics*, *113*(1).

Feng, X., Grossman, R., & Stein, L. (2011). PeakRanger: a cloud-enabled peak caller for ChIP-seq data. *BMC Bioinformatics*, *12*, 139.

Feng, S., Cokus, S. J., Schubert, V., Zhai, J., Pellegrini, M., & Jacobsen, S. E. (2014). Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. *Molecular Cell*, *55*(5), 694–707.

Ferdous, M., Higgins, J. D., Osman, K., Lambing, C., Roitinger, E., Mechtler, K., Armstrong, S. J., Perry, R., Pradillo, M., Cuñado, N., & Franklin, F. C. H. (2012). Inter-homolog crossing-over and synapsis in Arabidopsis meiosis are dependent on the chromosome axis protein AtASY3. *PLoS Genetics*, *8*(2), e1002507.

Fernandes, J. B., Séguéla-Arnaud, M., Larchevêque, C., Lloyd, A. H., & Mercier, R. (2018a). Unleashing meiotic crossovers in hybrid plants. *Proceedings of the National Academy of Sciences*, *115*(10), 2431–2436.

Fernandes, J. B., Duhamel, M., Seguéla-Arnaud, M., Froger, N., Girard, C., Choinard, S., Solier, V., De Winne, N., De Jaeger, G., Gevaert, K., Andrey, P., Grelon, M., Guerois, R., Kumar, R., & Mercier, R. (2018b). FIGL1 and its novel partner FLIP form a conserved complex that regulates homologous recombination. *PLOS Genetics*, *14*(4), e1007317.

Flores, O., & Orozco, M. (2011). nucleR: a package for non-parametric nucleosome positioning. *Bioinformatics*, *27*(15), 2149–2150.

Fowler, K. R., Sasaki, M., Milman, N., Keeney, S., and Smith, G. R. (2014). Evolutionarily diverse determinants of meiotic DNA break and recombination landscapes across the genome. *Genome Res.*, *24*(10), 1650–1664.

Francis, K. E., Lam, S. Y., Harrison, B. D., Bey, A. L., Berchowitz, L. E., & Copenhaver, G. P. (2007). Pollen tetrad-based visual assay for meiotic recombination in Arabidopsis. *Proceedings of the National Academy of Sciences*, *104*(10), 3913–3918.

Fransz, P., ten Hoopen, R., & Tessadori, F. (2006). Composition and formation of heterochromatin in *Arabidopsis thaliana*. *Chromosome Research*, *14*(1), 71–82.

Fransz, P., Linc, G., Lee, C.-R., Aflitos, S. A., Lasky, J. R., Toomajian, C., Ali, H., Peters, J., van Dam, P., Ji, X., Kuzak, M., Gerats, T., Schubert, I., Schneeberger, K., Colot, V., Martienssen, R., Koornneef, M., Nordborg, M., Juenger, T. E., de Jong, H., & Schranz, M. E. (2016). Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*. *The Plant Journal*, *88*(2), 159–178.

Frigola, J., Sabarinathan, R., Mularoni, L., Muiños, F., Gonzalez-Perez, A., & López-Bigas, N. (2017). Reduced mutation rate in exons due to differential mismatch repair. *Nature Genetics*, *49*(12), 1684–1692.

Fung, J. C., Rockmill, B., Odell, M., & Roeder, G. S. (2004). Imposition of Crossover Interference through the Nonrandom Distribution of Synapsis Initiation Complexes. *Cell*, *116*(6), 795–802.

Garcia, V., Phelps, S. E. L., Gray, S., & Neale, M. J. (2011). Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature*, *479*(7372), 241–244.

Garcia, V., Gray, S., Allison, R. M., Cooper, T. J., & Neale, M. J. (2015). Tel1ATM-mediated interference suppresses clustered meiotic double-strand-break formation. *Nature*, *520*(7545), 114–118.

Gari, K., Décaillet, C., Stasiak, A. Z., Stasiak, A., & Constantinou, A. (2008). The Fanconi Anemia Protein FANCM Can Promote Branch Migration of Holliday Junctions and Replication Forks. *Molecular Cell*, *29*(1), 141–148.

Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M. A., & Malinverni, R. (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics (Oxford, England)*, *32*(2), 289–291.

Genschel, J., Littman, S. J., Drummond, J. T., & Modrich, P. (1998). Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. *The Journal of Biological Chemistry*, *273*(31), 19895–19901.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., & Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature*, *425*(6959), 737–741.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345.

Girard, C., Crismani, W., Froger, N., Mazel, J., Lemhemdi, A., Horlow, C., & Mercier, R. (2014). FANCM-associated proteins MHF1 and MHF2, but not the other Fanconi anemia factors, limit meiotic crossovers. *Nucleic Acids Research*, *42*(14), 9087–9095.

Girard, C., Chelysheva, L., Choinard, S., Froger, N., Macaisne, N., Lehmemdi, A., Mazel, J., Crismani, W., & Mercier, R. (2015). AAA-ATPase FIDGETIN-LIKE 1 and Helicase FANCM Antagonize Meiotic Crossovers by Distinct Mechanisms. *PLOS Genetics*, *11*(7), e1005369.

Giraut, L., Falque, M., Drouaud, J., Pereira, L., Martin, O. C., & Mézard, C. (2011). Genome-Wide Crossover Distribution in *Arabidopsis thaliana* Meiosis Reveals Sex-Specific Patterns along Chromosomes. *PLoS Genetics*, *7*(11), e1002354.

Goldman, A. S., & Lichten, M. (2000). Restriction of ectopic recombination by interhomolog interactions during *Saccharomyces cerevisiae* meiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(17), 9537–9542.

Gorman, J., Chowdhury, A., Surtees, J. A., Shimada, J., Reichman, D. R., Alani, E., & Greene, E. C. (2007). Dynamic Basis for One-Dimensional DNA Scanning by the Mismatch Repair Complex Msh2-Msh6. *Molecular Cell*, *28*(3), 359–370.

Gouil, Q., & Baulcombe, D. C. (2016). DNA Methylation Signatures of the Plant Chromomethyltransferases. *PLOS Genetics*, *12*(12), e1006526.

Gradia, S., Acharya, S., & Fishel, R. (1997). The Human Mismatch Recognition Complex hMSH2-hMSH6 Functions as a Novel Molecular Switch. *Cell*, *91*(7), 995–1005.

Grilley, M., Welsh, K. M., Su, S. S., & Modrich, P. (1989). Isolation and characterization of the *Escherichia coli* mutL gene product. *The Journal of Biological Chemistry*, *264*(2), 1000–1004.

Groothuizen, F. S., Fish, A., Petoukhov, M. V., Reumer, A., Manelyte, L., Winterwerp, H. H., Marinus, M. G., Lebbink, J. H., Svergun, D. I., Friedhoff, P., & Sixma, T. K. (2013). Using stable MutS dimers and tetramers to quantitatively analyze DNA mismatch recognition and sliding clamp formation. *Nucleic Acids Research*, *41*(17), 8166–8181.

Groothuizen, F. S., & Sixma, T. K. (2016). The conserved molecular machinery in DNA mismatch repair enzyme structures. *DNA Repair*, *38*, 14–23.

Haber, L. T., Pang, P. P., Sobell, D. I., Mankovich, J. A., & Walker, G. C. (1988). Nucleotide sequence of the Salmonella typhimurium mutS gene required for mismatch repair: homology of MutS and HexA of Streptococcus pneumoniae. *Journal of Bacteriology*, *170*(1), 197–202.

Haracska, L., Unk, I., Johnson, R. E., Johansson, E., Burgers, P. M., Prakash, S., & Prakash, L. (2001). Roles of yeast DNA polymerases delta and zeta and of Rev1 in the bypass of abasic sites. *Genes & Development*, *15*(8), 945–954.

Hartung, F., Wurz-Wildersinn, R., Fuchs, J., Schubert, I., Suer, S., & Puchta, H. (2007). The Catalytically Active Tyrosine Residues of Both SPO11-1 and SPO11-2 Are Required for Meiotic Double-Strand Break Induction in *Arabidopsis*. *The Plant Cell*, *19*(10), 3090–3099.

Hauf, S., & Watanabe, Y. (2004). Kinetochore Orientation in Mitosis and Meiosis. *Cell*, *119*(3), 317–327.

Hays, J. B. (2002). *Arabidopsis thaliana*, a versatile model system for study of eukaryotic genome-maintenance functions. *DNA Repair*, *1*(8), 579–600.

He, Y., Wang, M., Dukowic-Schulze, S., Zhou, A., Tiang, C.-L., Shilo, S., Sidhu, G. K., Eichten, S., Bradbury, P., Springer, N. M., Buckler, E. S., Levy, A. A., Sun, Q., Pillardy, J., Kianian, P. M. A., Kianian, S. F., Chen, C., & Pawlowski, W. P. (2017). Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(46), 12231–12236.

He, W., Rao, H. B. D. P., Tang, S., Bhagwat, N., Kulkarni, D., Chang, M. A. W., Hall, C., Singh, L., Chen, X., Hollingsworth, N. M., Cejka, P., & Hunter, N. (2018). The crossover function of MutSγ is activated via Cdc7-dependent stabilization of Msh4. *BioRxiv*, 386458.

Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S., & Mullineaux, P. M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Molecular Biology*, *42*(6), 819–832.

Henikoff, S., & Henikoff, J. G. (2012). 'Point' centromeres of Saccharomyces harbor single centromere-specific nucleosomes. *Genetics*, *190*(4), 1575–1577.

Herman, G. E., & Modrich, P. (1981). *Escherichia coli* K-12 clones that overproduce dam methylase are hypermutable. *Journal of Bacteriology*, *145*(1), 644–646.

Heyer, W.-D., Ehmsen, K. T., & Liu, J. (2010). Regulation of Homologous Recombination in Eukaryotes. *Annual Review of Genetics*, *44*(1), 113–139.

Higgins, J. D., Armstrong, S. J., Franklin, F. C. H., & Jones, G. H. (2004). The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. *Genes & Development*, *18*(20), 2557–2570.

Higgins, J. D., Sanchez-Moran, E., Armstrong, S. J., Jones, G. H., & Franklin, F. C. H. (2005). The Arabidopsis synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. *Genes & Development*, *19*(20), 2488–2500.

Higgins, J. D., Buckling, E. F., Franklin, F. C. H., & Jones, G. H. (2008a). Expression and functional analysis of AtMUS81 in Arabidopsis meiosis reveals a role in the second pathway of crossing-over. *The Plant Journal*, *54*(1), 152–162.

Higgins, J. D., Vignard, J., Mercier, R., Pugh, A. G., Franklin, F. C. H., & Jones, G. H. (2008b). AtMSH5 partners AtMSH4 in the class I meiotic crossover pathway in *Arabidopsis thaliana*, but is not required for synapsis. *The Plant Journal*, *55*(1), 28–39.

Hilliker, A. J., Clark, S. H., & Chovnick, A. (1991). The effect of DNA sequence polymorphisms on intragenic recombination in the rosy locus of *Drosophila melanogaster*. *Genetics*, *129*(3).

Hoffman, P. D., Leonard, J. M., Lindberg, G. E., Bollmann, S. R., & Hays, J. B. (2004). Rapid accumulation of mutations during seed-to-seed propagation of mismatch-repair-defective Arabidopsis. *Genes & Development*, *18*(21), 2676–2685.

Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genetical Research*, *5*(02), 282.

Hollingsworth, N. M., Ponte, L., & Halsey, C. (1995). MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in *Saccharomyces cerevisiae* but not mismatch repair. *Genes & Development*, *9*(14), 1728–1739.

Hombauer, H., Campbell, C. S., Smith, C. E., Desai, A., & Kolodner, R. D. (2011). Visualization of Eukaryotic DNA Mismatch Repair Reveals Distinct Recognition and Repair Intermediates. *Cell*, *147*(5), 1040–1053.

Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, *16*(9), 907–916.

Horton, M. W., Hancock, A. M., Huang, Y. S., Toomajian, C., Atwell, S., Auton, A., Muliyati, N. W., Platt, A., Sperone, F. G., Vilhjálmsson, B. J., Nordborg, M., Borevitz, J. O., & Bergelson, J. (2012). Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics*, *44*(2), 212.

Hosouchi, T., Kumekawa, N., Tsuruoka, H., & Kotani, H. (2002). Physical Map-Based Sizes of the Centromeric Regions of *Arabidopsis thaliana* Chromosomes 1, 2, and 3. *DNA Research*, *9*(4), 117–121.

Huang, Y., Fang, J., Bedford, M. T., Zhang, Y., & Xu, R.-M. (2006). Recognition of Histone H3 Lysine-4 Methylation by the Double Tudor Domain of JMJD2A. *Science*, *312*(5774), 748–751.

Hunter, N., Chambers, S. R., Louis, E. J., & Borts, R. H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *The EMBO Journal*, *15*(7), 1726–1733.

Hunter, N. (2015). Meiotic Recombination: The Essence of Heredity. *Cold Spring Harbor Perspectives in Biology*, *7*(12), a016618.

International Wheat Genome Sequencing Consortium (IWGSC). (2014). A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science*, *345*(6194), 1251788–1251788.

Ito, H., Miura, A., Takashima, K., & Kakutani, T. (2006). Ecotype-specific and chromosome-specific expansion of variant centromeric satellites in *Arabidopsis thaliana*. *Molecular Genetics and Genomics*, *277*(1), 23–30.

Jackson, N., Sanchez-Moran, E., Buckling, E., Armstrong, S. J., Jones, G. H., & Franklin, F. C. H. (2006). Reduced meiotic crossovers and delayed prophase I progression in AtMLH3-deficient Arabidopsis. *The EMBO Journal*, *25*(6), 1315–1323.

Jacob, Y., Feng, S., LeBlanc, C. A., Bernatavichute, Y. V, Stroud, H., Cokus, S., Johnson, L. M., Pellegrini, M., Jacobsen, S. E., & Michaels, S. D. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nature Structural & Molecular Biology*, *16*(7), 763–768.

Javaid, S., Manohar, M., Punja, N., Mooney, A., Ottesen, J. J., Poirier, M. G., & Fishel, R. (2009). Nucleosome Remodeling by hMSH2-hMSH6. *Molecular Cell*, *36*(6), 1086–1094.

Jeffreys, A. J., Neumann, R., Panayi, M., Myers, S., & Donnelly, P. (2005). Human recombination hot spots hidden in regions of strong marker association. *Nature Genetics*, *37*(6), 601–606.

Jessop, L., Rockmill, B., Roeder, G. S., & Lichten, M. (2006). Meiotic Chromosome Synapsis-Promoting Proteins Antagonize the Anti-Crossover Activity of Sgs1. *PLoS Genetics*, *2*(9), e155.

Jiang, J., Bai, L., Surtees, J. A., Gemici, Z., Wang, M. D., & Alani, E. (2005). Detection of High-Affinity and Sliding Clamp Modes for MSH2-MSH6 by Single-Molecule Unzipping Force Analysis. *Molecular Cell*, *20*(5), 771–781.

Jiricny, J. (2013). Postreplicative Mismatch Repair. *Cold Spring Harbor Perspectives in Biology*, *5*(4), a012633–a012633.

Johnson, S. J., & Beese, L. S. (2004). Structures of mismatch replication errors observed in a DNA polymerase. *Cell*, *116*(6), 803–816.

Jones, G. H., & Franklin, F. C. H. (2006). Meiotic Crossing-over: Obligation and Interference. *Cell*, *126*(2), 246–248.

Kadyrov, F. A., Genschel, J., Fang, Y., Penland, E., Edelmann, W., & Modrich, P. (2009). A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *Proceedings of the National Academy of Sciences*, *106*(21), 8495–8500.

Kadyrova, L. Y., & Kadyrov, F. A. (2016). Endonuclease activities of MutLα and its homologs in DNA mismatch repair. *DNA Repair*, *38*, 42–49.

Karimova, G. A., Grigor'ev, P. S., & Rybchin, V. N. (1985). The role of genes of the system of mismatched base correction in genetic recombination of *Escherichia coli*. *Molekuliarnaia Genetika, Mikrobiologiia i Virusologiia*, (10), 29–34.

Kaufmann, K., Muiño, J. M., Østerås, M., Farinelli, L., Krajewski, P., & Angenent, G. C. (2010). Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nature Protocols*, *5*(3), 457–472.

Kauppi, L., May, C. A., & Jeffreys, A. J. (2009). Analysis of Meiotic Recombination Products from Human Sperm. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 557, pp. 323–355). Humana Press.

Kawakatsu, T., Huang, S. C., Jupe, F., Sasaki, E., Schmitz, R. J., Urich, M. A., Castanon, R., Nery, J. R., Barragan, C., He, Y., Chen, H., Dubin, M., Lee, C. R., Wang, C., Bemm, F., Becker, C., O'Neil, R., O'Malley, R. C., Quarless, D. X., 1001 Genomes Consortium, Schork, N. J., Weigel, D., Nordborg, M., & Ecker, J. R. (2016). Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cell*, *166*(2), 492–505.

Keeney, S., Giroux, C. N., & Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, *88*(3), 375–384.

Keeney, S., & Neale, M. J. (2006). Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochemical Society Transactions*, *34*(4), 523–525.

Kelmenson, P. M., Petkov, P., Wang, X., Higgins, D. C., Paigen, B. J., & Paigen, K. (2005). A torrid zone on mouse chromosome 1 containing a cluster of recombinational hotspots. *Genetics*, *169*(2), 833–841.

Kerzendorfer, C., Vignard, J., Pedrosa-Harand, A., Siwiec, T., Akimcheva, S., Jolivet, S., Sablowski, R., Armstrong, S., Schweizer, D., Mercier, R., & Schlögelhofer, P. (2006). The *Arabidopsis thaliana* MND1 homologue plays a key role in meiotic homologous pairing, synapsis and recombination. *Journal of Cell Science*, *119*(12), 2486–2496.

Kim, S., Plagnol, V., Hu, T. T., Toomajian, C., Clark, R. M., Ossowski, S., Ecker, J. R., Weigel, D., & Nordborg, M. (2007). Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, *39*(9), 1151–1155.

Kim, K. P., Weiner, B. M., Zhang, L., Jordan, A., Dekker, J., & Kleckner, N. (2010). Sister Cohesion and Structural Axis Components Mediate Homolog Bias of Meiotic Recombination. *Cell*, *143*(6), 924–937.

Kleckner, N. (2006). Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex. *Chromosoma*, *115*(3), 175–194.

Klein, F., Mahr, P., Galova, M., Buonomo, S. B. C., Michaelis, C., Nairz, K., & Nasmyth, K. (1999). A Central Role for Cohesins in Sister Chromatid Cohesion, Formation of Axial Elements, and Recombination during Yeast Meiosis. *Cell*, *98*(1), 91–103.

Kneitz, B., Cohen, P. E., Avdievich, E., Zhu, L., Kane, M. F., Hou, H., Kolodner, R. D., Kucherlapati, R., Pollard, J. W., Edelmann, W. (2000). MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. *Genes & Development*, *14*(9), 1085–1097.

Kohl, K. P., Jones, C. D., & Sekelsky, J. (2012). Evolution of an MCM complex in flies that promotes meiotic crossovers by blocking BLM helicase. *Science (New York, N.Y.)*, *338*(6112), 1363–1365.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S., & Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genetics*, *41*(3), 376–381.

Koornneef, M., and Meinke, D. (2004). The development of Arabidopsis as a model plant. *Plant Journal*, *61*(6), 909–921.

Kugou, K., Fukuda, T., Yamada, S., Ito, M., Sasanuma, H., Mori, S., Katou, Y., Itoh, T., Matsumoto, K., Shibata, T., Shirahige, K., & Ohta, K. (2009). Rec8 Guides Canonical Spo11 Distribution along Yeast Meiotic Chromosomes. *Molecular Biology of the Cell*, *20*(13), 3064–3076.

Kugou, K., & Ohta, K. (2009). Genome-Wide High-Resolution Chromatin Immunoprecipitation of Meiotic Chromosomal Proteins in *Saccharomyces cerevisiae* (pp. 285–304). Humana Press.

Kumar, C., Piacente, S. C., Sibert, J., Bukata, A. R., O'Connor, J., Alani, E., & Surtees, J. A. (2011). Multiple Factors Insulate Msh2–Msh6 Mismatch Repair Activity from Defects in Msh2 Domain I. *Journal of Molecular Biology*, *411*(4), 765–780.

Kunkel, T. A., & Erie, D. A. (2015). Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annual Review of Genetics 49*(1), 291–313.

Kunz, C., & Schär, P. (2004). Meiotic recombination: sealing the partnership at the junction. *Current Biology : CB*, *14*(22), R962-4.

Kurzbauer, M.-T., Uanschou, C., Chen, D., & Schlögelhofer, P. (2012). The Recombinases DMC1 and RAD51 Are Functionally and Spatially Separated during Meiosis in Arabidopsis. *The Plant Cell*, *24*(5), 2058–2070.

Laguri, C., Duband-Goulet, I., Friedrich, N., Axt, M., Belin, P., Callebaut, I., Gilquin, B., Zinn-Justin, S., & Couprie, J. (2008). Human Mismatch Repair Protein MSH6 Contains a PWWP Domain That Targets Double Stranded DNA. *Biochemistry*, *47*(23), 6199–6207.

Lahiri, S., & Mukerji, I. (2018). Homology Modeling and Structural Analysis of *S. Cerevisiae* Msh4 and Msh5 Provide Insight into DNA Binding and Specificity. *Biophysical Journal*, *114*(3), 85a.

Lahue, R. S., Su, S. S., & Modrich, P. (1987). Requirement for d(GATC) sequences in *Escherichia coli* mutHLS mismatch correction. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(6), 1482–1486.

Lahue, R. S., Au, K. G., & Modrich, P. (1989). DNA mismatch correction in a defined system. *Science (New York, N.Y.)*, *245*(4914), 160–164.

Lam, I., & Keeney, S. (2015). Mechanism and Regulation of Meiotic Recombination Initiation. *Cold Spring Harbor Perspectives in Biology*, *7*(1), a016634.

Lambing, C., Osman, K., Nuntasoontorn, K., West, A., Higgins, J. D., Copenhaver, G. P., Yang, J., Armstrong, S. J., Mechtler, K., Roitinger, E., & Franklin, F. C. H. (2015). Arabidopsis PCH2 Mediates Meiotic Chromosome Remodeling and Maturation of Crossovers. *PLOS Genetics*, *11*(7), e1005372.

Lambing, C., & Heckmann, S. (2018). Tackling Plant Meiosis: From Model Research to Crop Improvement. *Frontiers in Plant Science*, *9*, 829–829.

Lambing, C., Tock, A. J., Choi, K., Topp, S. D., Kuo, P. C., Blackwell, A. R., Zhao, X., Osman, K., Higgins, J. D., Franklin, F. C. H., & Henderson, I. R. (2019). REC8-cohesin, chromatin and transcription orchestrate meiotic recombination in the Arabidopsis genome. *BioRxiv*, 512400.

Lang, Z., Xie, S., & Zhu, J.-K. (2016). The 1001 Arabidopsis DNA Methylomes: An Important Resource for Studying Natural Genetic, Epigenetic, and Phenotypic Variation. *Trends in Plant Science*, *21*(11), 906–908.

Lange, J., Pan, J., Cole, F., Thelen, M. P., Jasin, M., & Keeney, S. (2011). ATM controls meiotic double-strand-break formation. *Nature*, *479*(7372), 237–240.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359.

LaRocque, J. R., & Jasin, M. (2010). Mechanisms of Recombination between Diverged Sequences in Wild-Type and BLM-Deficient Mouse and Human Cells. *Molecular and Cellular Biology*, *30*(8), 1887–1897.

Lawrence, K. S., & Engebrecht, J. (2012). Slowing Replication in Preparation for Reduction. *PLoS Genetics*, *8*(5), e1002715.

Lawrence, E. J., Griffin, C. H., & Henderson, I. R. (2017). Modification of meiotic recombination by natural variation in plants. *Journal of Experimental Botany*, *68*(20), 5471–5483.

Leonard, J. M., Bollmann, S. R., & Hays, J. B. (2003). Reduction of stability of arabidopsis genomic and transgenic DNA-repeat sequences (microsatellites) by inactivation of AtMSH2 mismatch-repair function., *133*(1).

León-Ortiz, A. M., Panier, S., Sarek, G., Vannier, J.-B., Patel, H., Campbell, P. J., & Boulton, S. J. (2018). A Distinct Class of Genome Rearrangements Driven by Heterologous Recombination. *Molecular Cell*, *69*(2), 292–305.e6.

Li, G. M., & Modrich, P. (1995). Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(6), 1950–1954.

Li, W., Chen, C., Markmann-Mulisch, U., Timofejeva, L., Schmelzer, E., Ma, H., & Reiss, B. (2004a). The Arabidopsis AtRAD51 gene is dispensable for vegetative development but required for meiosis. *Proceedings of the National Academy of Sciences*, *101*(29), 10596–10601.

Li, L., Santerre-Ayotte, S., Boivin, E. B., Jean, M., & Belzile, F. (2004b). A novel reporter for intrachromosomal homoeologous recombination in *Arabidopsis thaliana*. *The Plant Journal*, *40*(6), 1007–1015.

Li, W., Yang, X., Lin, Z., Timofejeva, L., Xiao, R., Makaroff, C. A., & Ma, H. (2005). The AtRAD51C gene is required for normal meiotic chromosome synapsis and double-stranded break repair in Arabidopsis. *Plant Physiology*, *138*(2), 965–976.

Li, L., Jean, M., & Belzile, F. (2006). The impact of sequence divergence and DNA mismatch repair on homeologous recombination in Arabidopsis. *The Plant Journal*, *45*(6), 908–916.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079.

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., & Li, G.-M. (2013). The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through Its Interaction with MutSα. *Cell*, *153*(3), 590–600.

Lippman, Z., May, B., Yordan, C., Singer, T., & Martienssen, R. (2003). Distinct Mechanisms Determine Transposon Inheritance and Methylation via Small Interfering RNA and Histone Modification. *PLoS Biology*, *1*(3), e67.

Liu, C., Wang, C., Wang, G., Becker, C., Zaidem, M., & Weigel, D. (2016). Genome-wide analysis of chromatin packing in *Arabidopsis thaliana* at single-gene resolution. *Genome Research*, *26*(8), 1057–1068.

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., Zhang, Q., Vilhjálmsson, B. J., Korte, A., Nizhynska, V., Voronin, V., Korte, P., Sedman, L., Mandáková, T., Lysak, M. A., Seren, Ü., Hellmann, I., & Nordborg, M. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics*, *45*(8), 884–890.

López, E., Pradillo, M., Oliver, C., Romero, C., Cun˜ado, N., & Santos, J. L. (2012). Looking for natural variation in chiasma frequency in *Arabidopsis thaliana*. *Journal of Experimental Botany*. Oxford University Press.

Lorenz, A., Osman, F., Sun, W., Nandi, S., Steinacher, R., & Whitby, M. C. (2012). The fission yeast FANCM ortholog directs non-crossover recombination during meiosis. *Science (New York, N.Y.)*, *336*(6088), 1585–1588.

Lu, X., Liu, X., An, L., Zhang, W., Sun, J., Pei, H., Meng, H., Fan, Y., & Zhang, C. (2008). The Arabidopsis MutS homolog AtMSH5 is required for normal meiosis. *Cell Research*, *18*(5), 589–599.

Lu, P., Han, X., Qi, J., Yang, J., Wijeratne, A. J., Li, T., & Ma, H. (2012). Analysis of Arabidopsis genome-wide variations before and after meiosis and meiotic recombination by resequencing Landsberg erecta and all four products of a single meiosis. *Genome Research*, *22*(3), 508–518.

Lujan, S. A., Clausen, A. R., Clark, A. B., MacAlpine, H. K., MacAlpine, D. M., Malc, E. P., Mieczkowski, P. A., Burkholder, A. B., Fargo, D. C., Gordenin, D. A., & Kunkel, T. A. (2014). Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Research*, *24*(11), 1751–1764.

Luo, Q., Tang, D., Wang, M., Luo, W., Zhang, L., Qin, B., Shen, Y., Wang, K., Li, Y., & Cheng, Z. (2013). The Role of OsMSH5 in Crossover Formation during Rice Meiosis. *Molecular Plant*, *6*(3), 729–742.

Ma, H. (2006). A molecular portrait of Arabidopsis meiosis. *The Arabidopsis Book*, *4*, e0095.

Macaisne, N., Novatchkova, M., Peirera, L., Vezon, D., Jolivet, S., Froger, N., Chelysheva, L., Grelon, M., & Mercier, R. (2008). SHOC1, an XPF Endonuclease-Related Protein, Is Essential for the Formation of Class I Meiotic Crossovers. *Current Biology*, *18*(18), 1432–1437.

Macaisne, N., Vignard, J., & Mercier, R. (2011). SHOC1 and PTD form an XPF-ERCC1-like complex that is required for formation of class I crossovers. *Journal of Cell Science*, *124*(Pt 16), 2687–2691.

MacQueen, A. J., & Roeder, G. S. (2009). Fpr3 and Zip3 Ensure that Initiation of Meiotic Recombination Precedes Chromosome Synapsis in Budding Yeast. *Current Biology*, *19*(18), 1519–1526.

Magni, G. E., & Von Borstel, R. C. (1962). Different Rates of Spontaneous Mutation during Mitosis and Meiosis in Yeast. *Genetics*, *47*(8), 1097–1108.

Mancera, E., Bourgon, R., Brozzi, A., Huber, W., & Steinmetz, L. M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature*, *454*(7203), 479–485.

Manhart, C. M., Ni, X., White, M. A., Ortega, J., Surtees, J. A., & Alani, E. (2017). The mismatch repair and meiotic recombination endonuclease Mlh1-Mlh3 is activated by polymer formation and can cleave DNA substrates in trans. *PLOS Biology*, *15*(4), e2001164.

Mankovich, J. A., McIntyre, C. A., & Walker, G. C. (1989). Nucleotide sequence of the Salmonella typhimurium mutL gene required for mismatch repair: homology of MutL to HexB of Streptococcus pneumoniae and to PMS1 of the yeast *Saccharomyces cerevisiae*. *Journal of Bacteriology*, *171*(10), 5325–5331.

Mannuss, A., Dukowic-Schulze, S., Suer, S., Hartung, F., Pacher, M., & Puchta, H. (2010). RAD5A, RECQ4A, and MUS81 have specific functions in homologous recombination and define different pathways of DNA repair in *Arabidopsis thaliana*. *The Plant Cell*, *22*(10), 3318–3330.

Marinus, M. G., Poteete, A., & Arraj, J. A. (1984). Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli* K-12. *Gene*, *28*(1), 123–125.

Marsischky, G. T., Lee, S., Griffith, J., & Kolodner, R. D. (1999). *'Saccharomyces cerevisiae* MSH2/6 complex interacts with Holliday junctions and facilitates their cleavage by phage resolution enzymes. *The Journal of Biological Chemistry*, *274*(11), 7200–7206.

Martini, E., Diaz, R. L., Hunter, N., & Keeney, S. (2006). Crossover Homeostasis in Yeast Meiosis. *Cell*, *126*(2), 285–295.

Martini, E., Borde, V., Legendre, M., Audic, S., Regnault, B., Soubigou, G., Dujon, B., & Llorente, B. (2011). Genome-Wide Analysis of Heteroduplex DNA in Mismatch Repair–Deficient Yeast Cells Reveals Novel Properties of Meiotic Recombination Pathways. *PLoS Genetics*, *7*(9), e1002305.

Mazina, O. M., Mazin, A. V, Nakagawa, T., Kolodner, R. D., & Kowalczykowski, S. C. (2004). *Saccharomyces cerevisiae* Mer3 Helicase Stimulates 3′–5′ Heteroduplex Extension by Rad51: Implications for Crossover Control in Meiotic Recombination. *Cell*, *117*(1), 47–56.

McCulloch, S. D., Kokoska, R. J., Chilkova, O., Welch, C. M., Johansson, E., Burgers, P. M. J., & Kunkel, T. A. (2004). Enzymatic switching for efficient and accurate translesion DNA replication. *Nucleic Acids Research*, *32*(15), 4665–4675.

McMahill, M. S., Sham, C. W., & Bishop, D. K. (2007). Synthesis-dependent strand annealing in meiosis. *PLoS Biology*, *5*(11), e299.

Meinke, D., Sweeney, C., & Muralla, R. (2009). Integrating the Genetic and Physical Maps of *Arabidopsis thaliana*: Identification of Mapped Alleles of Cloned Essential (EMB) Genes. *PLoS ONE*, *4*(10), e7386.

Méjean, V., & Claverys, J.-P. (1984). Effect of mismatched base pairs on the fate of donor DNA in transformation of Streptococcus pneumoniae. *MGG Molecular & General Genetics*, *197*(3), 467–471.

Melamed-Bessudo, C., Yehuda, E., Stuitje, A. R., & Levy, A. A. (2005). A new seed-based assay for meiotic recombination in *Arabidopsis thaliana*. *The Plant Journal*, *43*(3), 458–466.

Melamed-Bessudo, C., & Levy, A. A. (2012). Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. *Proceedings of the National Academy of Sciences*, *109*(16), E981–E988.

Mercier, R., Jolivet, S., Vezon, D., Huppe, E., Chelysheva, L., Giovanni, M., Nogué, F., Doutriaux, M. P., Horlow, C., Grelon, M., & Mézard, C. (2005). Two meiotic crossover classes cohabit in Arabidopsis: one is dependent on MER3, whereas the other one is not. *Current Biology*, *15*(8), 692–701.

Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N., & Grelon, M. (2015). The Molecular Biology of Meiosis in Plants. *Annual Review of Plant Biology*, *66*(1), 297–327.

Mézard, C., Tagliaro Jahns, M., & Grelon, M. (2015). Where to cross? New insights into the location of meiotic crossovers. *Trends in Genetics*, *31*(7), 393–401.

Mieulet, D., Aubert, G., Bres, C., Klein, A., Droc, G., Vieille, E., Rond-Coissieux, C., Sanchez, M., Dalmais, M, Mauxion, J. P., Rothan, C., Guiderdoni, E., & Mercier, R. (2018). Unleashing meiotic crossovers in crops. *Nature Plants*, *4*(12), 1010–1016.

Mirouze, M., Lieberman-Lazarovich, M., Aversano, R., Bucher, E., Nicolet, J., Reinders, J., & Paszkowski, J. (2012). Loss of DNA methylation affects the recombination landscape in Arabidopsis. *Proceedings of the National Academy of Sciences*, *109*(15), 5880–5885.

Mitchel, K., Zhang, H., Welz-Voegele, C., & Jinks-Robertson, S. (2010). Molecular Structures of Crossover and Noncrossover Intermediates during Gap Repair in Yeast: Implications for Recombination. *Molecular Cell*, *38*(2), 211–222.

Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.-M., Taneja, N., Folco, H. D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J., & Grewal, S. I. S. (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature*, *516*(7531), 432–435.

Modrich, P. (1987). DNA Mismatch Correction. *Annual Review of Biochemistry*, *56*(1), 435–466.

Modrich, P. (1991). Mechanisms and Biological Effects of Mismatch Repair. *Annual Review of Genetics*, *25*(1), 229–253.

Modrich, P. (2006). Mechanisms in Eukaryotic Mismatch Repair. *Journal of Biological Chemistry*, *281*(41), 30305–30309.

Modrich, P., & Lahue, R. (1996). Mismatch Repair in Replication Fidelity, Genetic Recombination, and Cancer Biology. *Annual Review of Biochemistry*, *65*(1), 101–133.

Munz, P. (1994). An analysis of interference in the fission yeast *Schizosaccharomyces pombe*. *Genetics*, *137*(3).

Myers, S., Bottolo, L., Freeman, C., McVean, G., & Donnelly, P. (2005). A Fine-Scale Map of Recombination Rates and Hotspots Across the Human Genome. *Science*, *310*(5746), 321–324.

Nakagawa, T., & Kolodner, R. D. (2002). *Saccharomyces cerevisiae* Mer3 is a DNA helicase involved in meiotic crossing over. *Molecular and Cellular Biology*, *22*(10), 3281–3291.

Nambiar, M., & Smith, G. R. (2018). Pericentromere-Specific Cohesin Complex Prevents Meiotic Pericentric DNA Double-Strand Breaks and Lethal Crossovers. *Molecular Cell*, *71*(4), 540–553.e4.

Neale, M. J., Ramachandran, M., Trelles-Sticken, E., Scherthan, H., & Goldman, A. S. H. (2002). Wild-type levels of Spo11-induced DSBs are required for normal single-strand resection during meiosis. *Molecular Cell*, *9*(4), 835–846.

Neale, M. J., Pan, J., & Keeney, S. (2005). Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature*, *436*(7053), 1053–1057.

Neyton, S., Lespinasse, F., Moens, P. B., Paul, R., Gaudray, P., Paquis-Flucklinger, V., & Santucci-Darmanin, S. (2004). Association between MSH4 (MutS homologue 4) and the DNA strand-exchange RAD51 and DMC1 proteins during mammalian meiosis. *MHR: Basic Science of Reproductive Medicine*, *10*(12), 917–924.

Nishant, K. T., Plys, A. J., & Alani, E. (2008). A Mutation in the Putative MLH3 Endonuclease Domain Confers a Defect in Both Mismatch Repair and Meiosis in *Saccharomyces cerevisiae*. *Genetics*, *179*(2), 747–755.

Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N. A., Shah, C., Wall, J. D., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M., & Bergelson, J. (2005). The Pattern of Polymorphism in *Arabidopsis thaliana*. *PLoS Biology*, *3*(7), e196.

Novak, J. E., Ross-Macdonald, P. B., & Roeder, G. S. (2001). The budding yeast Msh4 protein functions in chromosome synapsis and the regulation of crossover distribution. *Genetics*, *158*(3), 1013–1025.

Obmolova, G., Ban, C., Hsieh, P., & Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature*, *407*(6805), 703–710.

Ohkura, H. (2015). Meiosis: An Overview of Key Differences from Mitosis. *Cold Spring Harbor Perspectives in Biology*, *7*(5), a015859.

Oke, A., Anderson, C. M., Yam, P., & Fung, J. C. (2014). Controlling Meiotic Recombinational Repair – Specifying the Roles of ZMMs, Sgs1 and Mus81/Mms4 in Crossover Formation. *PLoS Genetics*, *10*(10), e1004690.

Opperman, R., Emmanuel, E., & Levy, A. A. (2004). The Effect of Sequence Divergence on Recombination Between Direct Repeats in Arabidopsis. *Genetics*, *168*(4), 2207–2215.

Osman, K., Sanchez-Moran, E., Mann, S. C., Jones, G. H., & Franklin, F. C. H. (2009). Replication protein A (AtRPA1a) is required for class I crossover formation but is dispensable for meiotic DNA break repair. *The EMBO Journal*, *28*(4), 394–404.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., Weigel, D., & Lynch, M. (2010). The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. *Science*, *327*(5961), 92–94.

Otto, S. P., & Gerstein, A. C. (2006). Why have sex? The population genetics of sex and recombination. *Biochemical Society Transactions*, *34*(4), 519–522.

Ozawa, K., Jergic, S., Park, A. Y., Dixon, N. E., & Otting, G. (2008). The proofreading exonuclease subunit epsilon of *Escherichia coli* DNA polymerase III is tethered to the polymerase subunit alpha via a flexible linker. *Nucleic Acids Research*, *36*(15), 5074–5082.

Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., & Jiricny, J. (1996). hMutSβ, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Current Biology*, *6*(9), 1181–1184.

Pan, J., Sasaki, M., Kniewel, R., Murakami, H., Blitzblau, H. G., Tischfield, S. E., Zhu, X., Neale, M. J., Jasin, M., Socci, N. D., Hochwagen, A., & Keeney, S. (2011). A Hierarchical Combination of Factors Shapes the Genome-wide Topography of Yeast Meiotic Recombination Initiation. *Cell*, *144*(5), 719–731.

Panizza, S., Mendoza, M. A., Berlinger, M., Huang, L., Nicolas, A., Shirahige, K., & Klein, F. (2011). Spo11-Accessory Proteins Link Double-Strand Break Sites to the Chromosome Axis in Early Meiotic Recombination. *Cell*, *146*(3), 372–383.

Paquis-Flucklinger, V., Santucci-Darmanin, S., Paul, R., Saunières, A., Turc-Carel, C., & Desnuelle, C. (1997). Cloning and Expression Analysis of a Meiosis-Specific MutS Homolog: The Human MSH4 Gene. *Genomics*, *44*(2), 188–194.

Pavlov, Y. I., Mian, I. M., & Kunkel, T. A. (2003). Evidence for Preferential Mismatch Repair of Lagging Strand DNA Replication Errors in Yeast. *Current Biology*, *13*(9), 744–748.

Perry, J., Kleckner, N., & Börner, G. V. (2005). Bioinformatic analyses implicate the collaborating meiotic crossover/chiasma proteins Zip2, Zip3, and Spo22/Zip4 in ubiquitin labeling. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(49), 17594–17599.

Picard, C. L., & Gehring, M. (2017). Proximal methylation features associated with nonrandom changes in gene body methylation. *Genome Biology*, *18*(1), 73.

Prieler, S., Penkner, A., Borde, V., & Klein, F. (2005). The control of Spo11's interaction with meiotic recombination hotspots. *Genes & Development*, *19*(2), 255–269.

Proffitt, J. H., Davie, J. R., Swinton, D., & Hattman, S. (1984). 5-Methylcytosine is not detectable in *Saccharomyces cerevisiae* DNA. *Molecular and Cellular Biology*, *4*(5), 985–988.

Puchta, H., & Hohn, B. (1991). The mechanism of extrachromosomal homologous DNA recombination in plant cells. *MGG Molecular & General Genetics*, *230*(1–2), 1–7.

Puchta, H., & Hohn, B. (2012). In planta somatic homologous recombination assay revisited: a successful and versatile, but delicate tool. *The Plant Cell*, *24*(11), 4324–4331.

Pukkila, P. J., Peterson, J., Herman, G., Modrich, P., & Meselson, M. (1983). Effects of high levels of dna adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. *Genetics*, *104*(4).

Qiao, H., Prasada Rao, H. B. D., Yang, Y., Fong, J. H., Cloutier, J. M., Deacon, D. C., Nagel, K. E., Swartz, R. K., Strong, E., Holloway, J. K., Cohen, P. E., Schimenti, J., Ward, J., & Hunter, N. (2014). Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination. *Nature Genetics*, *46*(2), 194–199.

Qin, S., & Min, J. (2014). Structure and function of the nucleosome-binding PWWP domain. *Trends in Biochemical Sciences*, *39*(11), 536–547.

Radman, M. (1989). Mismatch repair and the fidelity of genetic recombination. *Genome*, *31*(1), 68–73.

Rakshambikai, R., Srinivasan, N., & Nishant, K. T. (2013). Structural insights into *Saccharomyces cerevisiae* Msh4-Msh5 complex function using homology modeling. *PloS One*, *8*(11), e78753.

Rao, H. B. D. P., Qiao, H., Bhatt, S. K., Bailey, L. R. J., Tran, H. D., Bourne, S. L., Qiu, W., Deshpande, A., Sharma, A. N., Beebout, C. J., Pezza, R. J., & Hunter, N. (2017). A SUMO-ubiquitin relay recruits proteasomes to chromosome axes to regulate meiotic recombination. *Science (New York, N.Y.)*, *355*(6323), 403–407.

Rattray, A., Santoyo, G., Shafer, B., & Strathern, J. N. (2015). Elevated Mutation Rate during Meiosis in *Saccharomyces cerevisiae*. *PLoS Genetics*, *11*(1), e1004910.

Rayssiguier, C., Thaler, D. S., & Radman, M. (1989). The barrier to recombination between *Escherichia coli* and Salmonella typhimurium is disrupted in mismatch-repair mutants. *Nature*, *342*(6248), 396–401.

Reynolds, A., Qiao, H., Yang, Y., Chen, J. K., Jackson, N., Biswas, K., Holloway, J. K., Baudat, F., de Massy, B., Wang, J., Höög, C., Cohen, P. E., & Hunter, N. (2013). RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. *Nature Genetics*, *45*(3), 269–278.

Ribeiro, J., Abby, E., Livera, G., & Martini, E. (2016). RPA homologs and ssDNA processing during meiotic recombination. *Chromosoma*, *125*(2), 265–276.

Robert, T., Nore, A., Brun, C., Maffre, C., Crimi, B., Guichard, V., Bourbon, H.-M., & de Massy, B. (2016). The TopoVIB-Like protein family is required for meiotic DNA double-strand break formation. *Science*, *351*(6276), 943–949.

Rodgers-Melnick, E., Bradbury, P. J., Elshire, R. J., Glaubitz, J. C., Acharya, C. B., Mitchell, S. E., Li, C., Li, Y., & Buckler, E. S. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(12), 3823–3828.

Rogacheva, M. V, Manhart, C. M., Chen, C., Guarne, A., Surtees, J., & Alani, E. (2014). Mlh1-Mlh3, a meiotic crossover and DNA mismatch repair factor, is a Msh2-Msh3-stimulated endonuclease. *The Journal of Biological Chemistry*, *289*(9), 5664–5673.

Romanienko, P. J., & Camerini-Otero, R. D. (2000). The mouse Spo11 gene is required for meiotic chromosome synapsis. *Molecular Cell*, *6*(5), 975–987.

Romanova, N. V., & Crouse, G. F. (2013). Different Roles of Eukaryotic MutS and MutL Complexes in Repair of Small Insertion and Deletion Loops in Yeast. *PLoS Genetics*, *9*(10), e1003920.

Ross, K. J., Fransz, P., & Jones, G. H. (1996). A light microscopic atlas of meiosis in *Arabidopsis thaliana*. *Chromosome Research : An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *4*(7), 507–516.

Ross-Macdonald, P., & Roeder, G. S. (1994). Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. *Cell*, *79*(6), 1069–1080.

Rowan, B. A., Patel, V., Weigel, D., & Schneeberger, K. (2015). Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. *G3 (Bethesda, Md.)*, *5*(3), 385–398.

Sabouri, N., Viberg, J., Goyal, D. K., Johansson, E., & Chabes, A. (2008). Evidence for lesion bypass by yeast replicative DNA polymerases during DNA damage. *Nucleic Acids Research*, *36*(17), 5660–5667.

Salomé, P. A., Bomblies, K., Fitz, J., Laitinen, R. A. E., Warthmann, N., Yant, L., & Weigel, D. (2012). The recombination landscape in *Arabidopsis thaliana* F2 populations. *Heredity*, *108*(4), 447–455.

Sanchez-Moran, E., Santos, J.-L., Jones, G. H., & Franklin, F. C. H. (2007). ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in Arabidopsis. *Genes & Development*, *21*(17), 2220–2233.

Santucci-Darmanin, S., Neyton, S., Lespinasse, F., Saunières, A., Gaudray, P., & Paquis-Flucklinger, V. (2002). The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination. *Human Molecular Genetics*, *11*(15), 1697–1706.

Saparbaev, M., Prakash, L., & Prakash, S. (1996). Requirement of Mismatch Repair Genes MSH2 and MSH3 in the RAD1-RAD10 Pathway of Mitotic Recombination in *Saccharomyces cerevisiae*. *Genetics*, *142*(3).

Sarma, S., Pandey, A. K., Sharma, K., Ravi, M., Sreelakshmi, Y., & Sharma, R. (2018). MutS-Homolog2 silencing generates tetraploid meiocytes in tomato (*Solanum lycopersicum*). *Plant Direct*, *2*(1), e00017.

Schalbetter, S. A., Fudenberg, G., Baxter, J., Pollard, K. S., & Neale, M. J. (2018). Principles of Meiotic Chromosome Assembly. *BioRxiv*, 442038.

Schröpfer, S., Kobbe, D., Hartung, F., Knoll, A., & Puchta, H. (2014). Defining the roles of the N-terminal region and the helicase activity of RECQ4A in DNA repair and homologous recombination in Arabidopsis. *Nucleic Acids Research*, *42*(3), 1684–1697.

Schwartz, S., Meshorer, E., & Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nature Structural & Molecular Biology*, *16*(9), 990–995.

Séguéla-Arnaud, M., Crismani, W., Larchevêque, C., Mazel, J., Froger, N., Choinard, S., Lemhemdi, A., Macaisne, N., Van Leene, J., Gevaert, K., De Jaeger, G., Chelysheva, L., & Mercier, R. (2015). Multiple mechanisms limit meiotic crossovers: TOP3α and two BLM homologs antagonize crossovers in parallel to FANCM. *Proceedings of the National Academy of Sciences*, *112*(15), 4713–4718.

Séguéla-Arnaud, M., Choinard, S., Larchevêque, C., Girard, C., Froger, N., Crismani, W., & Mercier, R. (2017). RMI1 and TOP3α limit meiotic CO formation through their C-terminal domains. *Nucleic Acids Research*, *45*(4).

Selva, E. M., Maderazo, A. B., & Lahue, R. S. (1997). Differential effects of the mismatch repair genes MSH2 and MSH3 on homeologous recombination in *Saccharomyces cerevisiae*. *MGG - Molecular & General Genetics*, *257*(1), 71–82.

Serra, H., Lambing, C., Griffin, C. H., Topp, S. D., Nageswaran, D. C., Underwood, C. J., Ziolkowski, P. A., Séguéla-Arnaud, M., Fernandes, J. B., Mercier, R., & Henderson, I. R. (2018a). Massive crossover elevation via combination of HEI10 and recq4a recq4b during Arabidopsis meiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(10), 2437–2442.

Serra, H., Choi, K., Zhao, X., Blackwell, A. R., Kim, J., & Henderson, I. R. (2018b). Interhomolog polymorphism shapes meiotic crossover within the Arabidopsis *RAC1* and *RPP13* disease resistance genes. *PLoS Genetics*, *14*(12).

Shen, P., & Huang, H. V. (1986). Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics*, *112*(3).

Shen, P., & Huang, H. V. (1989). Effect of base pair mismatches on recombination via the RecBCD pathway. *MGG Molecular & General Genetics*, *218*(2), 358–360.

Shen, Y., Tang, D., Wang, K., Wang, M., Huang, J., Luo, W., Luo, Q., Hong, L., Li, M., & Cheng, Z. (2012). ZIP4 in homologous chromosome synapsis and crossover formation in rice meiosis. *Journal of Cell Science*, *125*(Pt 11), 2581–2591.

Shibata, A., Moiani, D., Arvai, A. S., Perry, J., Harding, S. M., Genois, M.-M., Maity, R., van Rossum-Fikkert, S., Kertokalio, A., Romoli, F., Ismail, A., Ismalaj, E., Petricci, E., Neale, M. J., Bristow, R. G., Masson, J. Y., Wyman, C., Jeggo, P. A., & Tainer, J. A. (2014). DNA Double-Strand Break Repair Pathway Choice Is Directed by Distinct MRE11 Nuclease Activities. *Molecular Cell*, *53*(1), 7–18.

Shibutani, S., Takeshita, M., & Grollman, A. P. (1991). Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature*, *349*(6308), 431–434.

Shilo, S., Melamed-Bessudo, C., Dorone, Y., Barkai, N., & Levy, A. A. (2015). DNA Crossover Motifs Associated with Epigenetic Modifications Delineate Open Chromatin Regions in Arabidopsis. *The Plant Cell*, *27*(9), 2427–2436.

Shinohara, M., Hayashihara, K., Grubb, J. T., Bishop, D. K., & Shinohara, A. (2015). DNA damage response clamp 9-1-1 promotes assembly of ZMM proteins for formation of crossovers and synaptonemal complex. *Journal of Cell Science*, *128*(8), 1494–1506.

Shultz, R. W., Tatineni, V. M., Hanley-Bowdoin, L., & Thompson, W. F. (2007). Genome-wide analysis of the core DNA replication machinery in the higher plants Arabidopsis and rice. *Plant Physiology*, *144*(4), 1697–1714.

Sia, E. A., Kokoska, R. J., Dominska, M., Greenwell, P., & Petes, T. D. (1997). Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Molecular and Cellular Biology*, *17*(5), 2851–2858.

Simoens, C. R., Gielen, J., Van Montagu, M., & Inzé, D. (1988). Characterization of highly repetitive sequences of *Arabidopsis thaliana*. *Nucleic Acids Research*, *16*(14), 6753–6766.

Singhal, S., Leffler, E. M., Sannareddy, K., Turner, I., Venn, O., Hooper, D. M., Strand, A. I., Li, Q., Raney, B., Balakrishnan, C. N., Griffith, S. C., McVean, G., & Przeworski, M. (2015). Stable recombination hotspots in birds. *Science*, *350*(6263), 928–932.

Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, *8*(4), 272–285.

Smagulova, F., Gregoretti, I. V., Brick, K., Khil, P., Camerini-Otero, R. D., & Petukhova, G. V. (2011). Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature*, *472*(7343), 375–378.

Smyth, D. R., Bowman, J. L., & Meyerowitz, E. M. (1990). Early flower development in Arabidopsis. *The Plant Cell*, *2*(8), 755–767.

Snowden, T., Acharya, S., Butz, C., Berardini, M., & Fishel, R. (2004). hMSH4-hMSH5 Recognizes Holliday Junctions and Forms a Meiosis-Specific Sliding Clamp that Embraces Homologous Chromosomes. *Molecular Cell*, *15*(3), 437–451.

Snowden, T., Shim, K.-S., Schmutte, C., Acharya, S., & Fishel, R. (2008). hMSH4-hMSH5 adenosine nucleotide processing and interactions with homologous recombination machinery. *The Journal of Biological Chemistry*, *283*(1), 145–154.

Sommermeyer, V., Béneut, C., Chaplais, E., Serrentino, M. E., & Borde, V. (2013). Spp1, a Member of the Set1 Complex, Promotes Meiotic DSB Formation in Promoters by Tethering Histone H3K4 Methylation Sites to Chromosome Axes. *Molecular Cell*, *49*(1), 43–54.

Stacey, N. J., Kuromori, T., Azumi, Y., Roberts, G., Breuer, C., Wada, T., Maxwell, A., Roberts, K., & Sugimoto-Shirasu, K. (2006). Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. *The Plant Journal*, *48*(2), 206–216.

Stanzione, M., Baumann, M., Papanikos, F., Dereli, I., Lange, J., Ramlal, A., Tränkner, D., Shibuya, H., de Massy, B., Watanabe, Y., Jasin, M., Keeney, S., & Tóth, A. (2016). Meiotic DNA break formation requires the unsynapsed chromosome axis-binding protein IHO1 (CCDC36) in mice. *Nature Cell Biology*, *18*(11), 1208–1220.

Storlazzi, A., Gargano, S., Ruprich-Robert, G., Falque, M., David, M., Kleckner, N., & Zickler, D. (2010). Recombination Proteins Mediate Meiotic Spatial Chromosome Organization and Pairing. *Cell*, *141*(1), 94–106.

Strathern, J. N., Shafer, B. K., & McGill, C. B. (1995). DNA synthesis errors associated with double-strand-break repair. *Genetics*, *140*(3).

Stroud, H., Greenberg, M. V. C., Feng, S., Bernatavichute, Y. V, & Jacobsen, S. E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell*, *152*(1–2), 352–364.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D. J., & Jacobsen, S. E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nature Structural & Molecular Biology*, *21*(1), 64–72.

Su, S. S., & Modrich, P. (1986). *Escherichia coli* mutS-encoded protein binds to mismatched DNA base pairs. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(14), 5057–5061.

Su, S. S., Lahue, R. S., Au, K. G., & Modrich, P. (1988). Mispair specificity of methyl-directed DNA mismatch correction in vitro. *The Journal of Biological Chemistry*, *263*(14), 6829–6835.

Su, S.-S., Grilley, M., Thresher, R., Griffith, J., & Modrich, P. (1989). Gap formation is associated with methyl-directed mismatch correction under conditions of restricted DNA synthesis. *Genome*, *31*(1), 104–111.

Sun, Y., Ambrose, J. H., Haughey, B. S., Webster, T. D., Pierrie, S. N., Muñoz, D. F., Wellman, E. C., Cherian, S., Lewis, S. M., Berchowitz, L. E., & Copenhaver, G. P. (2012). Deep Genome-Wide Measurement of Meiotic Gene Conversion Using Tetrad Analysis in *Arabidopsis thaliana*. *PLoS Genetics*, *8*(10), e1002968.

Sun, X., Huang, L., Markowitz, T. E., Blitzblau, H. G., Chen, D., Klein, F., & Hochwagen, A. (2015). Transcription dynamically patterns the meiotic chromosome-axis interface. *ELife*, *4*.

Sun, H., Rowan, B. A., Flood, P. J., Brandt, R., Fuss, J., Hancock, A. M., Michelmore, R. W., Huettel, B., & Schneeberger, K. (2018). Linked-read sequencing of gametes allows efficient genome-wide analysis of meiotic recombination. *BioRxiv*, 484022.

Sym, M., Engebrecht, J. A., & Roeder, G. S. (1993). ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. *Cell*, *72*(3), 365–378.

Syrjänen, J. L., Pellegrini, L., & Davies, O. R. (2014). A molecular model for the role of SYCP3 in meiotic chromosome organisation. *ELife*, *3*.

Szakasits, D., Siddique, S., & Bohlmann, H. (2007). An Improved pPZP Vector for Agrobacterium-mediated Plant Transformation. *Plant Molecular Biology Reporter*, *25*(3–4), 115–120.

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., & Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell*, *33*(1), 25–35.

Tanaka, T., Cosma, M. P., Wirth, K., & Nasmyth, K. (1999). Identification of Cohesin Association Sites at Centromeres and along Chromosome Arms. *Cell*, *98*(6), 847–858.

Terasawa, M., Ogawa, H., Tsukamoto, Y., Shinohara, M., Shirahige, K., Kleckner, N., & Ogawa, T. (2007). Meiotic recombination-related DNA synthesis and its implications for cross-over and non-cross-over recombinant formation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(14), 5965–5970.

Terui, R., Nagao, K., Kawasoe, Y., Taki, K., Higashi, T. L., Tanaka, S., Nakagawa, T., Obuse, C., Masukata, H., & Takahashi, T. S. (2018). Nucleosomes around a mismatched base pair are excluded via an Msh2-dependent reaction with the aid of SNF2 family ATPase Smarcad1. *Genes & Development*, *32*(11–12), 806–821.

Tessadori, F., van Zanten, M., Pavlova, P., Clifton, R., Pontvianne, F., Snoek, L. B., Millenaar, F. F., Schulkes, R. K., van Driel, R., Voesenek, L. A., Spillane, C., Pikaard, C. S., Fransz, P., & Peeters, A. J. M. (2009). PHYTOCHROME B and HISTONE DEACETYLASE 6 Control Light-Induced Chromatin Compaction in *Arabidopsis thaliana*. *PLoS Genetics*, *5*(9), e1000638.

Thacker, D., Mohibullah, N., Zhu, X., & Keeney, S. (2014). Homologue engagement controls meiotic DNA break number and distribution. *Nature*, *510*(7504), 241–246.

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, *408*(6814), 796–815.

Thomas, D. C., Roberts, J. D., & Kunkel, T. A. (1991). Heteroduplex repair in extracts of human HeLa cells. *The Journal of Biological Chemistry*, *266*(6), 3744–3751.

Tock, A. J., & Henderson, I. R. (2018). Hotspots for Initiation of Meiotic Recombination. *Frontiers in Genetics*, *9*, 521.

Trouiller, B., Schaefer, D. G., Charlot, F., & Nogué, F. (2006). MSH2 is essential for the preservation of genome integrity and prevents homeologous recombination in the moss Physcomitrella patens. *Nucleic Acids Research*, *34*(1), 232–242.

Uanschou, C., Siwiec, T., Pedrosa-Harand, A., Kerzendorfer, C., Sanchez-Moran, E., Novatchkova, M., … Schlögelhofer, P. (2007). A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. *The EMBO Journal*, *26*(24), 5061–5070.

Underwood, C. J., Henderson, I. R., & Martienssen, R. A. (2017). Genetic and epigenetic variation of transposable elements in Arabidopsis. *Current Opinion in Plant Biology*, *36*, 135–141.

Underwood, C. J., Choi, K., Lambing, C., Zhao, X., Serra, H., Borges, F., Simorowski, J., Ernst, E., Jacob, Y., Henderson, I. R., & Martienssen, R. A. (2018). Epigenetic activation of meiotic recombination near *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. *Genome Research*, *28*(4), 519–531.

Urig, S., Gowher, H., Hermann, A., Beck, C., Fatemi, M., Humeny, A., & Jeltsch, A. (2002). The *Escherichia coli* Dam DNA Methyltransferase Modifies DNA in a Highly Processive Reaction. *Journal of Molecular Biology*, *319*(5), 1085–1096.

Van Marcke, I., & Angenon, G. (2013). Genomic stability in Nicotiana plants upon silencing of the mismatch repair gene MSH2. *Plant Biotechnology Reports*, *7*(4), 467–480.

Voelkel-Meiman, K., Johnston, C., Thappeta, Y., Subramanian, V. V, Hochwagen, A., & MacQueen, A. J. (2015). Separable Crossover-Promoting and Crossover-Constraining Aspects of Zip1 Activity during Budding Yeast Meiosis. *PLoS Genetics*, *11*(6), e1005335.

Vrielynck, N., Chambon, A., Vezon, D., Pereira, L., Chelysheva, L., De Muyt, A., Mézard, C., Mayer, C., & Grelon, M. (2016). A DNA topoisomerase VI-like complex initiates meiotic recombination. *Science*, *351*(6276), 939–943.

Vulić, M., Dionisio, F., Taddei, F., & Radman, M. (1997). Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(18), 9763–9767.

Wagner, R., & Meselson, M. (1976). Repair tracts in mismatched DNA heteroduplexes. *Proceedings of the National Academy of Sciences of the United States of America*, *73*(11), 4135–4139.

Wang, H., & Hays, J. B. (2004). Signaling from DNA mispairs to mismatch-repair excision sites despite intervening blockades. *The EMBO Journal*, *23*(10), 2126–2133.

Wang, K., Wang, M., Tang, D., Shen, Y., Miao, C., Hu, Q., Lu, T., & Cheng, Z. (2012). The Role of Rice HEI10 in the Formation of Meiotic Crossovers. *PLoS Genetics*, *8*(7), e1002809.

Wang, C., Wang, Y., Cheng, Z., Zhao, Z., Chen, J., Sheng, P., Yu, Y., Ma, W., Duan, E., Wu, F., Liu, L., Qin, R., Zhang, X., Guo, X., Wang, J., Jiang, L., & Wan, J. (2016). The role of OsMSH4 in male and female gamete development in rice meiosis. *Journal of Experimental Botany*, *67*(5), 1447–1459.

Wang, Y., & Copenhaver, G. P. (2018). Meiotic Recombination: Mixing It Up in Plants. *Annual Review of Plant Biology*, *69*(1), 577–609.

Warburton, P. E., Giordano, J., Cheung, F., Gelfand, Y., & Benson, G. (2004). Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Research*, *14*(10A), 1861–1869.

Watanabe, Y., & Nurse, P. (1999). Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature*, *400*(6743), 461–464.

Watson, J. M., Platzer, A., Kazda, A., Akimcheva, S., Valuchova, S., Nizhynska, V., Nordborg, M., & Riha, K. (2016). Germline replications and somatic mutation accumulation are independent of vegetative life span in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(43), 12226–12231.

Watt, V. M., Ingles, C. J., Urdea, M. S., & Rutter, W. J. (1985). Homology requirements for recombination in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(14), 4768–4772.

Weigel, D. (2012). Natural variation in Arabidopsis: from molecular genetics to ecological genomics. *Plant Physiology*, *158*(1), 2–22.

Welz-Voegele, C., & Jinks-Robertson, S. (2008). Sequence Divergence Impedes Crossover More Than Noncrossover Events During Mitotic Gap Repair in Yeast. *Genetics*, *179*(3), 1251–1262.

Wijeratne, A. J., Zhang, W., Sun, Y., Liu, W., Albert, R., Zheng, Z., Oppenheimer, D. G., Zhao, D., & Ma, H. (2007). Differential gene expression in Arabidopsis wild-type and mutant anthers: insights into anther cell differentiation and regulatory networks. *The Plant Journal*, *52*(1), 14–29.

Wijnker, E., Velikkakam James, G., Ding, J., Becker, F., Klasen, J. R., Rawat, V., Rowan, B. A., de Jong, D. F., de Snoo, C. B., Zapata, L., Huettel, B., de Jong, H., Ossowski, S., Weigel, D., Koornneef, M., Keurentjes, J. J., & Schneeberger, K. (2013). The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *ELife*, *2*, e01426.

Worth, L., Clark, S., Radman, M., & Modrich, P. (1994). Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(8), 3238–3241.

Wu, S.-Y., Culligan, K., Lamers, M., & Hays, J. (2003). Dissimilar mispair-recognition spectra of Arabidopsis DNA-mismatch-repair proteins MSH2.MSH6 (MutS) and MSH2.MSH7 (MutS). *Nucleic Acids Research*, *31*(20), 6027–6034.

Wu, G., Rossidivito, G., Hu, T., Berlyand, Y., & Poethig, R. S. (2015). Traffic Lines: New Tools for Genetic Analysis in *Arabidopsis thaliana*. *Genetics*, *200*(1), 35–45.

Xu, H., Beasley, M. D., Warren, W. D., van der Horst, G. T. J., & McKay, M. J. (2005). Absence of Mouse REC8 Cohesin Promotes Synapsis of Sister Chromatids in Meiosis. *Developmental Cell*, *8*(6), 949–961.

Xue, M., Wang, J., Jiang, L., Wang, M., Wolfe, S., Pawlowski, W. P., Wang, Y., & He, Y. (2018). The Number of Meiotic Double-Strand Breaks Influences Crossover Distribution in Arabidopsis. *The Plant Cell*, *30*(10), 2628–2638.

Yamaguchi, M., Dao, V., & Modrich, P. (1998). MutS and MutL activate DNA helicase II in a mismatch-dependent manner. *The Journal of Biological Chemistry*, *273*(15), 9197–9201.

Yang, F., De La Fuente, R., Leu, N. A., Baumann, C., McLaughlin, K. J., & Wang, P. J. (2006). Mouse SYCP2 is required for synaptonemal complex assembly and chromosomal synapsis during male meiosis. *The Journal of Cell Biology*, *173*(4), 497–507.

Yang, S., Yuan, Y., Wang, L., Li, J., Wang, W., Liu, H., Chen, J. Q., Hurst, L. D., & Tian, D. (2012). Great majority of recombination events in Arabidopsis are gene conversion events. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(51), 20992–20997.

Yelagandula, R., Stroud, H., Holec, S., Zhou, K., Feng, S., Zhong, X., Muthurajan, U. M., Nie, X., Kawashima, T., Groth, M., Luger, K., Jacobsen, S. E., & Berger, F. (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. *Cell*, *158*(1), 98–109.

Yelina, N. E., Choi, K., Chelysheva, L., Macaulay, M., de Snoo, B., Wijnker, E., Miller, N., Drouaud, J., Grelon, M., Copenhaver, G. P., Mezard, C., Kelly, K. A., & Henderson, I. R. (2012). Epigenetic Remodeling of Meiotic Crossover Frequency in *Arabidopsis thaliana* DNA Methyltransferase Mutants. *PLoS Genetics*, *8*(8), e1002844.

Yelina, N. E., Ziolkowski, P. A., Miller, N., Zhao, X., Kelly, K. A., Muñoz, D. F., Mann, D. J., Copenhaver, G. P., & Henderson, I. R. (2013). High-throughput analysis of meiotic crossover frequency and interference via flow cytometry of fluorescent pollen in *Arabidopsis thaliana*. *Nature Protocols*, *8*(11), 2119–2134.

Yelina, N. E., Lambing, C., Hardcastle, T. J., Zhao, X., Santos, B., & Henderson, I. R. (2015). DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis., *29*(20).

Yoshida, K., Kondoh, G., Matsuda, Y., Habu, T., Nishimune, Y., & Morita, T. (1998). The mouse RecA-like gene Dmc1 is required for homologous chromosome synapsis during meiosis. *Molecular Cell*, *1*(5), 707–718.

Zakharyevich, K., Tang, S., Ma, Y., & Hunter, N. (2012). Delineation of Joint Molecule Resolution Pathways in Meiosis Identifies a Crossover-Specific Resolvase. *Cell*, *149*(2), 334–347.

Zalevsky, J., MacQueen, A. J., Duffy, J. B., Kemphues, K. J., Villeneuve, A. M., Li, Y., & Cheng, Z. (1999). Crossing over during *Caenorhabditis elegans* meiosis requires a conserved MutS-based pathway that is partially dispensable in budding yeast. *Genetics*, *153*(3), 1271–1283.

Zapata, L., Ding, J., Willing, E.-M., Hartwig, B., Bezdan, D., Jiao, W.-B., Patel, V., Velikkakam James, G., Koornneef, M., Ossowski, S., & Schneeberger, K. (2016). Chromosome-level assembly of *Arabidopsis thaliana* Ler reveals the extent of translocation and inversion polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(28), E4052-60.

Zetka, M. C., & Rose, A. M. (1995). Mutant rec-1 eliminates the meiotic pattern of crossing over in *Caenorhabditis elegans*. *Genetics*, *141*(4), 1339–1349.

Zhang, Y., Yuan, F., Presnell, S. R., Tian, K., Gao, Y., Tomkinson, A. E., Gu, L., & Li, G.-M. (2005). Reconstitution of 5′-Directed Human Mismatch Repair in a Purified System. *Cell*, *122*(5), 693–705.

Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., & Chua, N.-H. (2006). Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nature Protocols*, *1*(2), 641–646.

Zhang, X., Bernatavichute, Y. V, Cokus, S., Pellegrini, M., & Jacobsen, S. E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biology*, *10*(6), R62.

Zhang, L., Kim, K. P., Kleckner, N. E., & Storlazzi, A. (2011a). Meiotic double-strand breaks occur once per pair of (sister) chromatids and, via Mec1/ATR and Tel1/ATM, once per quartet of chromatids. *Proceedings of the National Academy of Sciences*, *108*(50), 20036–20041.

Zhang, L., Ma, H., & Pugh, B. F. (2011b). Stable and dynamic nucleosome states during a meiotic developmental process. *Genome Research*, *21*(6), 875–884.

Zhang, C., Song, Y., Cheng, Z., Wang, Y., Zhu, J., Ma, H., Xu., L., & Yang, Z.-N. (2012). The *Arabidopsis thaliana* DSB formation (*AtDFO*) gene is required for meiotic double-strand break formation. *The Plant Journal*, *72*(2), 271–281.

Zhang, L., Tang, D., Luo, Q., Chen, X., Wang, H., Li, Y., & Cheng, Z. (2014). Crossover Formation During Rice Meiosis Relies on Interaction of OsMSH4 and OsMSH5. *Genetics*, *198*(4), 1447–1456.

Zhang, T., Zhang, W., & Jiang, J. (2015). Genome-Wide Nucleosome Occupancy and Positioning and Their Impact on Gene Expression and Evolution in Plants. *Plant Physiology*, *168*(4), 1406–1416.

Zhao, S., Zhang, B., Yang, M., Zhu, J., & Li, H. (2018). Systematic Profiling of Histone Readers in *Arabidopsis thaliana*. *Cell Reports*, *22*(4), 1090–1102.

Zheng, G. X. Y., Lau, B. T., Schnall-Levin, M., Jarosz, M., Bell, J. M., Hindson, C. M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D. A., Merrill, L., Terry, J. M., Mudivarti, P. A., Wyatt, P. W., Bharadwaj, R., Makarewicz, A. J., Li, Y., Belgrader, P., Price, A. D., Lowe, A. J., Marks, P., Vurens, G. M., Hardenbol, P., Montesclaros, L., Luo, M., Greenfield, L., Wong, A., Birch, D. E., Short, S. W., Bjornson, K. P., Patel, P., Hopmans, E. S., Wood, C., Kaur, S., Lockwood, G. K., Stafford, D., Delaney, J. P., Wu, I., Ordonez, H. S., Grimes, S. M., Greer, S., Lee, J. Y., Belhocine, K., Giorda, K. M., Heaton, W. H., McDermott, G. P., Bent, Z. W., Meschi, F., Kondov, N. O., Wilson, R., Bernate, J. A., Gauby, S., Kindwall, A., Bermejo, C., Fehr, A. N., Chan, A., Saxonov, S., Ness, K. D., Hindson, B. J., & Ji, H. P. (2016). Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature Biotechnology*, *34*(3), 303–311.

Zickler, D., & Kleckner, N. (1999). Meiotic Chromosomes: Integrating Structure and Function. *Annual Review of Genetics*, *33*(1), 603–754.

Zickler, D., & Kleckner, N. (2015). Recombination, Pairing, and Synapsis of Homologs during Meiosis. *Cold Spring Harbor Perspectives in Biology*, *7*(6), a016626.

Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. *Genome Biology*, *18*(1), 87.

Ziolkowski, P. A., Berchowitz, L. E., Lambing, C., Yelina, N. E., Zhao, X., Kelly, K. A., Choi, K., Ziolkowska, L., June, V., Sanchez-Moran, E., Franklin, F. C. H., Copenhaver, G. P., & Henderson, I. R. (2015). Juxtaposition of heterozygous and homozygous regions causes reciprocal crossover remodelling via interference during Arabidopsis meiosis. *ELife*, *4*.

Ziolkowski, P. A., & Henderson, I. R. (2017). Interconnections between meiotic recombination and sequence polymorphism in plant genomes. *New Phytologist*, *213*(3), 1022–1029.

Ziolkowski, P. A., Underwood, C. J., Lambing, C., Martinez-Garcia, M., Lawrence, E. J., Ziolkowska, L., Griffin, C., Choi, K., Franklin, F. C. H., Martienssen, R. A., & Henderson, I. R. (2017). Natural variation and dosage of the HEI10 meiotic E3 ligase control *Arabidopsis* crossover recombination. *Genes & Development*, *31*(3), 306–317.

Zmienko, A., Samelak-Czajka, A., Kozlowski, P., Szymanska, M., & Figlerowicz, M. (2016). *Arabidopsis thaliana* population analysis reveals high plasticity of the genomic region spanning MSH2, AT3G18530 and AT3G18535 genes and provides evidence for NAHR-driven recurrent CNV events occurring in this location. *BMC Genomics*, *17*(1), 893.